



HAL
open science

Expressões proverbiais do português

Sónia Margarida Moreira Reis

► **To cite this version:**

Sónia Margarida Moreira Reis. Expressões proverbiais do português. Linguistics. Universidade do Algarve, 2020. Portuguese. NNT: . tel-03849024

HAL Id: tel-03849024

<https://hal.science/tel-03849024v1>

Submitted on 11 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SÓNIA MARGARIDA MOREIRA REIS

EXPRESSÕES PROVERBIAIS DO PORTUGUÊS
Usos, Variação Formal e Identificação Automática

12 de outubro de 2020

Júri:
Pedro Alfonso Ferré da Ponte
Dolors Català Guitart
Manuel Célio Jesus Conceição
Jorge Manuel Evangelista Baptista
Cláudia Susana Nunes Martins



UNIVERSIDADE DO ALGARVE
Faculdade de Ciências Humanas e Sociais
2020

SÓNIA MARGARIDA MOREIRA REIS

EXPRESSÕES PROVERBIAIS DO PORTUGUÊS
Usos, Variação Formal e Identificação Automática

Doutoramento em Ciências da Linguagem

Trabalho efetuado sob a orientação de:
Prof. Doutor Jorge Manuel Evangelista Baptista



UNIVERSIDADE DO ALGARVE
Faculdade de Ciências Humanas e Sociais
2020

EXPRESSÕES PROVERBIAIS DO PORTUGUÊS

Usos, Variação Formal e Identificação Automática

Declaração de autoria de trabalho

Declaro ser a autora deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

Sónia Margarida Moreira Reis

(assinatura)

Copyright

Copyright: Sónia Margarida Moreira Reis.

A Universidade do Algarve reserva para si o direito, em conformidade com o disposto no Código do Direito de Autor e dos Direitos Conexos, de arquivar, reproduzir e publicar a obra, independentemente do meio utilizado, bem como de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição para fins meramente educacionais ou de investigação e não comerciais, conquanto seja dado o devido crédito ao autor e editor respetivos.

Agradecimentos

A realização desta tese contou com o importante contributo de algumas pessoas, às quais gostaria de expressar a minha profunda gratidão e afeto.

Ao Professor Doutor Jorge Baptista, pela total disponibilidade e confiança depositada e pelo privilégio da sua orientação e apoio. Muito obrigada por estimular a minha vontade constante de saber, pelos saberes que me transmitiu, pelas opiniões e críticas e por todas as palavras de incentivo.

Ao Professor Doutor Nuno Mamede, pelas contribuições que fez ao longo desta pesquisa.

À Universidade do Algarve, pelo auxílio financeiro concedido para a participação e apresentação em conferências internacionais de dois dos artigos desta tese.

Aos ilustres professores da Universidade do Algarve por tudo quanto me ensinaram.

A todos aqueles que disponibilizaram o seu tempo para responderem aos questionários de provérbios desta tese.

Quero também agradecer aos membros da L2F, Annamaria Pompili, Alberto Abad e Luísa Coheur pelo acesso à plataforma VITHEA e pelo apoio prestado.

Gostaria ainda de agradecer aos meus familiares e amigos por me terem acompanhado nesta jornada e pela tolerância e compreensão.

Ao Rogério, pela companhia e pelo apoio incondicional.

Um agradecimento especial à Vera por me ter incentivado a fazer este curso de Doutoramento.

Muito obrigada!

Resumo

Os provérbios são uma expressão da cultura de uma sociedade e estão ligados às mais diversas áreas da experiência humana. Este tipo de expressões surge nos mais variados tipos de texto e desempenha diferentes funções retóricas no discurso, nele se integrando por meio de diferentes mecanismos, nem sempre fáceis de detetar formalmente. A interferência nos processos de coerência e de coesão discursivas, p. ex., a referência, e a sua variação formal, constituem sérios desafios ao processamento da linguagem natural (PLN), exigindo a sua identificação e delimitação precisas.

Este projeto visa a identificação automática de provérbios portugueses (e as suas variantes) em textos, a fim de melhor caracterizar a sua utilização, tanto qualitativa como quantitativamente.

Tal permitirá a definição de índices de frequência e, a partir destes, bem como de outros critérios, a determinação da disponibilidade lexical das unidades paremiológicas (o provérbio e as suas variantes).

Estas informações são elementos relevantes a considerar, por exemplo, no desenvolvimento de instrumentos complementares de diagnóstico ou terapia de determinadas patologias da linguagem; ou mesmo para a construção de jogos didáticos para o ensino-aprendizagem de português, língua materna e língua não materna, eventualmente assistido por computador.

Palavras-chave: provérbios, variação formal, identificação automática, índices de frequência, disponibilidade lexical.

Abstract

Proverbs are an expression of the culture of a society and are connected to the most diverse areas of human experience. This type of expression occurs in the most diverse types of text and performs different rhetorical functions in the discourse, integrating it through several mechanisms, not always easy to detect formally. Interference in the processes of discursive coherence and cohesion, e.g. reference, and its formal variation, constitute serious challenges to Natural Language Processing (NLP), requiring its precise identification and delimitation.

This project aims at the automatic identification of Portuguese proverbs (and their variants) in texts, to effectively characterize their use, both qualitatively and quantitatively.

This will allow the definition of frequency indices and, from these, as well as other criteria, the determination of the lexical availability of the paremiological units (the proverb and its variants).

This information is relevant elements to consider, for example, in the development of complementary diagnostic or therapeutic instruments for certain language pathologies; or even for the construction of didactic games for the teaching-learning of Portuguese, either as a Mother Tongue or as a Foreign Language, eventually assisted by computer.

Keywords: proverbs, formal variation, automatic identification, frequency indices, lexical availability.

Índice

Lista de Abreviaturas	xv
Lista de Figuras	xvii
Lista de Tabelas	xxiii
I	1
Introdução	3
1.1 Objeto de estudo	5
1.1.1 O provérbio enquanto unidade de descrição linguística	5
1.1.2 A unidade provérbio e outros tipos de unidade no sistema linguístico	7
1.1.3 Uma visão tradicional	8
1.1.4 O provérbio: Marginalidade e centralidade na língua	9
1.1.5 Outra perspetiva	10
1.1.6 Provérbios e tradução	11
1.1.7 Uma visão integradora da dimensão discursiva dos provérbios: Zuluaga (1980)	13
1.1.8 Contribuição do modelo do Léxico-Gramática	16
1.1.9 Provérbios e Genericidade	17
1.1.10 Estrutura das frases proverbiais	18
1.1.11 Provérbios e unidades paremiológicas (UP)	19
1.1.12 Variação formal e classificação sintática	21
1.1.13 Representação formal das unidades paremiológicas (UP)	22
1.1.14 O mínimo paremiológico (MP)	22
1.1.15 Utilização do MP para aplicações didáticas	23
1.1.16 Utilização do MP para aplicações terapêuticas/diagnóstico	23
1.1.17 Relação do provérbio com o texto	24
1.1.18 Funções discursivas	27
1.1.19 Objetivos e métodos	29

1.1.20 Estrutura da tese	33
II Publicações	35
1 Portuguese Proverbs: Types and Variants	39
2 Let's play with proverbs? NLP tools and resources for iCALL applications around proverbs for PFL	51
3 O uso de provérbios no ensino de português	73
4 Os provérbios em manuais de ensino de português língua não materna	95
5 Estimating Lexical Availability of European Portuguese Proverbs	107
6 O provérbio como estímulo num terapeuta virtual	123
7 Estudo das estruturas proverbiais com sujeito indefinido: Provérbio ou expressão fixa?	145
8 Descrição e classificação sintática das expressões proverbiais do português europeu	167
9 Determinação de um mínimo paremiológico do português europeu	181
10 As intenções comunicativas no uso de provérbios	199
III	217
Conclusão	218
Referências Bibliográficas	231
Anexos	245
.1 Lista de <i>stopwords</i>	245
.2 Critérios para agrupar variantes de provérbios sob uma mesma <i>Unidade</i> <i>Paremiológica</i>	247
.3 Avaliação do OCR (<i>Optical Character Recognition</i>)	249
.4 Grafos	251
.5 Questionários	299

.6	Mínimo paremiológico do português europeu	381
.7	100 jogos com provérbios com disponibilidade lexical	389
.8	Jogos VITHEA–Kids	403
.9	Seleção provérbios usuais – Anotador 1	409
.10	Seleção provérbios usuais – Anotador 2	419

Lista de Abreviaturas

AVC	Acidente Vascular Cerebral (<i>Cerebrovascular Accident</i>)
EF	Expressão Fixa
FST	Finite State Transducer (<i>Transdutor de Estados Finitos</i>)
iCALL	Intelligent Computer-Assisted Language Learning (<i>Ensino de Línguas Assistido por Computador</i>)
MP	Mínimo Paremiológico
PE	Português Europeu (<i>European Portuguese</i>)
PEL	Portfolio Europeu das Línguas (ELP <i>European Language Portfolio</i>)
PLE	Português Língua Estrangeira (PFL <i>Portuguese as a Foreign Language</i>)
PLM	Português Língua Materna (<i>Portuguese Mother Tongue</i>)
PLN	Processamento da Linguagem Natural (NLP <i>Natural Language Processing</i>)
PLNM	Português Língua não Materna (PFL <i>Portuguese as a Foreign Language</i>)
QECR	Quadro Europeu Comum de Referência para as Línguas (CEFR <i>Common European Framework of Reference for Languages</i>)
RST	Rhetorical Structure Theory (<i>Teoria da Estrutura Retórica</i>)
UP	Unidade Paremiológica

Lista de Figuras

3.1	Figuras 3 e 4 a cores	73
10.1	Saída (parcial) do sistema STRING para um exemplo retirado da revista <i>Máxima</i>	230
2	Grafo de inserção 1_3	251
3	Grafo de inserção 1_5	251
4	Grafo de pontuação e <E>	251
5	Grafo: A beleza está nos olhos de que a vê	252
6	Grafo: A carne é fraca	252
7	Grafo: A culpa morreu solteira	252
8	Grafo: A curiosidade matou o gato	252
9	Grafo: Para baixo todos os santos ajudam	253
10	Grafo: A esperança é sempre a última a morrer	253
11	Grafo: A fome é o melhor tempero	253
12	Grafo: A intenção é que conta	254
13	Grafo: A justiça tarda mas não falta	254
14	Grafo: A experiência é a mãe da ciência	254
15	Grafo: A falar é que a gente se entende	254
16	Grafo: A cavalo dado não se olha ao dente	255
17	Grafo: A laranja pela manhã é oiro, ao meio-dia prata, à tarde cobre e à noite mata	255
18	Grafo: A mentira tem pernas curtas	255
19	Grafo: À noite todos os gatos são pardos	255
20	Grafo: A ocasião faz o ladrão	255
21	Grafo: A pensar morreu um burro	256
22	Grafo: Roma e Pavia não se fizeram num dia	256
23	Grafo: A preguiça morreu à sede à beira do rio	256
24	Grafo: A mulher honrada não tem ouvidos	256
25	Grafo: Quem tem unhas é que toca guitarra	257
26	Grafo: Não se deve cuspir no prato onde se comeu	257

27	Grafo: A fé move montanhas	257
28	Grafo: Não se apanham trutas a bragas enxutas	258
29	Grafo: Todos os caminhos vão dar a Roma	258
30	Grafo: A galinha da vizinha é sempre melhor que a minha	259
31	Grafo: Quem sabe, sabe	259
32	Grafo: Quem vai para o mar avia-se em terra	259
33	Grafo: Há sempre uma primeira vez para tudo	259
34	Grafo: Zangam-se as comadres, descobrem-se as verdades	260
35	Grafo: Nem só de pão vive o homem	260
36	Grafo: Recordar é viver	260
37	Grafo: Depois da tempestade vem a bonança	261
38	Grafo: Não há duas sem três	261
39	Grafo: Um mal nunca vem só	261
40	Grafo: Quem não tem dinheiro não tem vícios	262
41	Grafo: São mais as vozes do que as nozes	262
42	Grafo: O tempo voa	262
43	Grafo: Boa asa com todo o tempo voa	262
44	Grafo: A morte não escolhe idades	263
45	Grafo: A paciência tem limites	263
46	Grafo: A preguiça é a mãe de todos os vícios	263
47	Grafo: A pressa é inimiga da perfeição	263
48	Grafo: A roupa suja lava-se em casa	264
49	Grafo: A saúde não tem preço	264
50	Grafo: A união faz a força	264
51	Grafo: A verdade sempre vem à tona	264
52	Grafo: À primeira cavadela minhoca	265
53	Grafo: A vida são dois dias	265
54	Grafo: A vingança serve-se fria	265
55	Grafo: A voz do povo é a voz de Deus	265
56	Grafo: As árvores morrem de pé	265
57	Grafo: Quem conta um conto acrescenta um ponto	266
58	Grafo: Um crime não justifica outro	266
59	Grafo: Da discussão nasce a luz	266
60	Grafo: Quem está mal que se mude	266
61	Grafo: Há gente para tudo	267
62	Grafo: Há males que vêm por bem	267
63	Grafo: Na hora do aperto é que se conhecem os amigos	267
64	Grafo: As aparências iludem	268
65	Grafo: A montanha pariu um rato	268

66	Grafo: Um santo não pode estar em dois altares	268
67	Grafo: Diz-me com quem andas dir-te-ei quem és	268
68	Grafo: Quem vê caras não vê corações	269
69	Grafo: Mais vale tarde do que nunca	269
70	Grafo: Santos da casa não fazem milagres	269
71	Grafo: Quem não chora não mama	269
72	Grafo: A idade não perdoa	270
73	Grafo: O hábito não faz o monge	270
74	Grafo: Não há rosas sem espinhos	270
75	Grafo: Deus não dorme	271
76	Grafo: Não devas a quem deveu nem sirvas a quem serviu	271
77	Grafo: Não faças aos outros o que não gostas que te façam a ti	271
78	Grafo: Amor ausente amor para sempre	271
79	Grafo: Antes se perca a lã que a ovelha	272
80	Grafo: Mais vale prevenir que remediar	272
81	Grafo: Nunca sirvas a quem serviu nem peças a quem pediu	272
82	Grafo: Nunca peças a quem pediu nem devas a quem deveu	272
83	Grafo: Pensa duas vezes antes de falar	273
84	Grafo: Querer é poder	273
85	Grafo: Não sujes a água que hás de beber	273
86	Grafo: No tempo quente refresca o ventre	273
87	Grafo: Os homens não se medem aos palmos	273
88	Grafo: O que arde cura	274
89	Grafo: Quem sai aos seus não degenera	274
90	Grafo: Deitar cedo e cedo erguer dá saúde e faz crescer	274
91	Grafo: Antes dobrar que quebrar	274
92	Grafo: O fruto proibido é sempre o mais apetecido	275
93	Grafo: Gordura é formosura	275
94	Grafo: Morra Marta morra farta	275
95	Grafo: O seguro morreu de velho	275
96	Grafo: Casa onde não há pão todos ralham e ninguém tem razão	276
97	Grafo: Um homem prevenido vale por dois	276
98	Grafo: Quem mais fala mais erra	276
99	Grafo: Tempo é dinheiro	277
100	Grafo: Abril águas mil	277
101	Grafo: Água mole em pedra dura tanto dá até que fura	277
102	Grafo: Águas passadas não movem moinhos	278
103	Grafo: Albarda-se o burro à vontade do dono	278
104	Grafo: Amanhã é outro dia	279

105	Grafo: Amanhã também é dia	279
106	Grafo: Amigo não empata amigo	279
107	Grafo: Amigos, amigos, negócios à parte	279
108	Grafo: Amor com amor se paga	279
109	Grafo: Ano novo, vida nova	279
110	Grafo: Aprender até morrer	280
111	Grafo: A má ação fica com quem a pratica	280
112	Grafo: Mais vale só que mal acompanhado	280
113	Grafo: O que é achado não é roubado	281
114	Grafo: Hoje por mim amanhã por ti	281
115	Grafo: Hoje meu amanhã teu	281
116	Grafo: O sol quando nasce é para todos	281
117	Grafo: Para vilão, vilão e meio	282
118	Grafo: Quem te avisa teu amigo é	282
119	Grafo: Com o tempo tudo se cura	282
120	Grafo: A água o dá a água o leva	283
121	Grafo: Ao teu amigo não encubras segredo	283
122	Grafo: Na arca aberta o justo peca	283
123	Grafo: O bem soa o mal voa	284
124	Grafo: Boas contas fazem bons amigos	284
125	Grafo: Bom vinho bom vinagre	284
126	Grafo: Cão que ladra não morde	284
127	Grafo: A cão mordido todos o mordem	285
128	Grafo: A boda e a mortalha no céu se talha	285
129	Grafo: Deus nos livre de justiças novas e de chaminés velhas	285
130	Grafo: Quem se quer bem sempre se encontra	285
131	Grafo: Cada um com o seu ofício	286
132	Grafo: É bem casada a que não tem sogra nem cunhada	286
133	Grafo: O que é meu é meu o que é teu é nosso	286
134	Grafo: Quem não é por mim é contra mim	286
135	Grafo: A homem ruivo e mulher barbuda de longe os saúda	287
136	Grafo: Entre a honra e o dinheiro o segundo é o primeiro	287
137	Grafo: Quem o seu inimigo poupa nas mãos lhe morre	287
138	Grafo: O leitão de um mês, o pato de três	288
139	Grafo: A quem medo hão o seu logo lhe dão	288
140	Grafo: Da tua mulher e do amigo esperto não creias senão o que souberes ao certo	288
141	Grafo: A mulher ainda que rica seja se é pedida mais deseja	288
142	Grafo: Não perde venda senão quem não tem que venda	289

143	Grafo: Quem pés não tem coices não promete	289
144	Grafo: Porfia mata veado e não besteiro cansado	289
145	Grafo: Queijo do Alentejo, vinho de Lamego	289
146	Grafo: Quem o rabo corta por trás se descobre	290
147	Grafo: Em terra de cegos quem tem um olho é rei	290
148	Grafo: Quem não usa não cuida	290
149	Grafo: Velho que se si cuida cem anos dura	291
150	Grafo: O vinho e o amigo o mais antigo	291
151	Grafo: Ao amigo que não é certo, um olho fechado e outro aberto . . .	291
152	Grafo: Para bom entendedor meia palavra basta	292
153	Grafo: Quem não te conhece que te compre	292
154	Grafo: Gota a gota o mar se esgota	292
155	Grafo: O que a mulher quer Deus quer	293
156	Grafo: Pobre bispo, pobre serviço	293
157	Grafo: Quem para si não sabe, não ponha escola	293
158	Grafo: Antes da sopa, molha-se a boca	293
159	Grafo: Quanto tens, quanto vales	294
160	Grafo: Quem não arrisca não petisca	294
161	Grafo: Quem muito faz ou arrisca um dia o demo o petisca	294
162	Grafo: Cada um sabe as linhas com que se cose	294
163	Grafo: Quem canta não assobia	295
164	Grafo: Quem não se farta a comer não se farta a lamber	295
165	Grafo: Pela linha vem a tinha	295
166	Grafo: A quem muda Deus o ajuda	295
167	Grafo: A mulher e a sardinha querem-se da mais pequenina	296
168	Grafo: A noite é boa conselheira	296
169	Grafo: O homem na praça e a mulher em casa	296
170	Grafo: A uma hora cai a casa	296
171	Grafo: Com melão vinho de tostão	297
172	Grafo: No minguante de janeiro, corta o madeiro	297
173	Grafo: Quem de nada duvida nada sabe	297
174	Grafo: Quem outrem serve não é livre	297
175	Grafo: A pequeno passarinho pequeno ninho	298
176	Grafo: Quem pode pode	298
177	Grafo: Quem pode e não quer quando quer não pode	298
178	Grafo: Não há bela sem senão	298

Lista de Tabelas

10.1	Classes formais dos provérbios do português europeu (Adaptado de Reis & Baptista 2019b)	225
2	Avaliação OCR	249
3	100 jogos com provérbios com disponibilidade lexical	391
4	Jogos VITHEA–Kids	405
5	Seleção provérbios usuais – Anotador 1	409
6	Seleção provérbios usuais – Anotador 2	419

Parte I

Introdução

A frase seguinte, retirada do jornal *Público*¹

Costa defende que todos devem “meter a colher” contra a violência sobre mulheres²

é interessante por várias razões. Em primeiro lugar, ela é interpretável como uma expressão fixa, idiomática (*alguém <meter> a colher* (ou *colherada*) *em alguma coisa*), que poderia ser definida como “interferir” (Neves 2000: 277); “intervir”, “[I]ntrometer-se em conversa a que não se é chamado” (Simões 2000: 435); ou “Intrometer-se no que não lhe diz respeito” (Priberam)³. Disso é sinal o emprego das aspas, que alertam o leitor para o caráter não composicional da expressão. Podem ser complementos de *meter a colher* (*em*) variadíssimos nomes de diferentes subclasses semânticas como, por exemplo, nomes abstratos, [negritos nossos]

Há uma área da Capital Europeia da Cultura invisível a olho nu, mas que mete a colher em **grande parte dos espectáculos que compõem a programação do evento**⁴

Até a CGTP meteu a colher no **assunto**, tendo o seu líder pedindo (em público e dirigindo-se a Seguro), que “pense duas vezes” sobre as negociações com PSD e CDS, afirmando que “não se pode salvar aquilo que não merece ser salvo”⁵

Jorge Mendes ‘mete a colher’ no **passo de Fábio Silva**⁶

ou mesmo nomes que exprimam coletivos humanos, metonimicamente interpretados como agentes de um processo:

Didier Drogba, o homem que faz a diferença no ataque, meteu a colher na **defesa**⁷

¹<https://www.publico.pt/> (Último acesso: 11 de janeiro de 2020; todos os restantes url referidos neste documento foram verificados nesta data.)

²<https://www.publico.pt/2019/11/25/politica/noticia/costa-defende-meter-colher-violencia-mulheres-1894958>

³“colher”, in Dicionário Priberam da Língua Portuguesa [em linha], 2008-2013, <https://dicionario.priberam.org/colher> [consultado em 18-12-2019].

⁴<https://www.publico.pt/2001/03/25/jornal/equipa-de-producao-da-porto-2001-faz-um-pouco-de-tudo-155986>

⁵<https://www.psd.pt/wp-content/uploads/povolivre/ficheiro1389625649.pdf>

⁶<https://desporto.sapo.pt/futebol/primeira-liga/artigos/jorge-mendes-mete-a-colher-no-passe-de-fabio-silva>

⁷<https://www.jn.pt/arquivo/2005/jose-mourinho-furioso-com-ricardo-carvalho-508428.html>

“Rodeado” de jornalistas e em ambiente empresarial, Morais Sarmiento
“meteu a colher” no **mundo dos média**⁸

Num segundo momento, a referência a *mulheres* (e num segundo plano, a evocação de *violência*) remete o leitor para uma expressão proverbial, reconhecida da maioria dos falantes:

Entre marido e mulher nunca metas a colher

Este provérbio, entendido como expressão da sabedoria popular, recomenda/adverte que não se interfira na vida doméstica de um casal, pois os problemas conjugais deverão ser resolvidos entre o próprio casal.

Não é evidente, no entanto, os limites do conjunto de expressões proverbiais “reconhecidas pela maioria dos falantes” da língua. Efetivamente, as coletâneas de provérbios já publicadas reúnem não só – como é natural – os provérbios usuais, mas também grande número de outras expressões, de uso marginal, literário ou mesmo raro, quando não mesmo expressões claramente arcaicas e provavelmente caídas em desuso, apenas (re)copiadas de coletânea para coletânea. Efetivamente, não é uma tarefa fácil determinar o “núcleo duro” dos provérbios usuais da língua. Pelo contrário, tal implica uma definição clara das fronteiras de um provérbio, o desenvolvimento de uma metodologia de recolha e análise, que permita, nomeadamente, a aferição da frequência de emprego de um dado provérbio na linguagem contemporânea; bem como a clarificação do estatuto teórico do provérbio, por exemplo, enquanto elemento do léxico de uma língua (ou outro estatuto que for mais adequado).

A existência tanto do provérbio como da expressão idiomática levanta várias questões linguísticas interessantes. Sem se cair na falsa questão da prioridade ontológica entre “o ovo” e “a galinha”, é de destacar que a *colher* é provavelmente mais o resultado dos mecanismos mnemónicos (rima, métrica, etc.) que cristalizam a forma do provérbio, parecendo apontar para a prioridade deste sobre a expressão fixa. Ainda assim, a liberdade distribucional da expressão fixa sugere o tema dos processos de formação das expressões idiomáticas e dos provérbios e as complexas relações que entre os dois tipos de expressão se podem estabelecer.

Finalmente, note-se como a apropriação, por A. Costa, do provérbio e da expressão idiomática faz sobressair o tema da reutilização criativa/expressiva deste tipo de objeto linguístico. Contudo, ainda que qualquer falante seja capaz de estabelecer a relação entre o que A. Costa disse e as duas expressões – o provérbio, uma frase exclusivamente negativa; e a expressão fixa *meter a colher em*, expressão neutra em relação à forma de frase (afirmativa ou negativa); não é evidente que esse processo possa ser facilmente

⁸<https://omirante.pt/semanario/2003-12-25/economia/2003-12-24-ministro-pede-noticias-pela-positiva>

replicado por sistemas de processamento de linguagem natural, no seu estado atual de desenvolvimento.

Esta tese aborda algumas das problemáticas acima levantadas por este discurso. O formato da tese segue o modelo de *coletânea de artigos* (Parte II), precedidos de uma Introdução (Parte I), para enquadramento da investigação realizada, e seguidos de uma terceira parte (Parte III), onde se realçam os aspetos mais importantes desenvolvidos ao longo deste processo, as principais conclusões a que chegámos, bem como os resultados obtidos e as contribuições deste estudo.

Ainda que cada artigo contextualize os conteúdos tratados e descreva a metodologia utilizada, esta introdução apresentará alguns assuntos, de carácter mais geral, que julgámos poderem contribuir para uma melhor perceção global do trabalho desenvolvido. Assim, começamos por tentar delimitar o objeto de estudo deste trabalho: *os provérbios do português europeu*, (Secção §1.1) tentando ao mesmo tempo determinar o lugar dos provérbios na gramática da língua, relacionando-os quer com as expressões idiomáticas (ou frases fixas) quer com outras formas de fixidez combinatória que se manifestam na língua. Revisitaremos, nesse passo, as perspetivas apresentadas por alguns gramáticos, bem como as diferentes posições teóricas de vários linguistas ao longo do tempo, sem, contudo, ter procurado fazer uma revisão teórica exaustiva.

1.1 Objeto de estudo

Começamos por tentar delimitar o objeto de estudo deste trabalho: *os provérbios do português europeu*. Embora possa parecer simples, este não é, como veremos um objeto fácil de determinar.

1.1.1 O provérbio enquanto unidade de descrição linguística

Como em muitas outras disciplinas, a determinação das unidades do léxico não é uma tarefa trivial. Por um lado, é fácil concordar que há, no léxico, unidades mais facilmente identificáveis, formadas por uma só palavra gráfica e com um significado e construção característicos: são as *palavras simples*, que podem ser organizadas em diferentes classes gramaticais ou *classes de palavras*, de natureza morfossintática e semântica (v.g., nomes, adjetivos, advérbios, conjunções, preposições, etc.). Por outro lado, há certas combinações de palavras simples que, no seu conjunto, também formam novas unidades lexicais: são as *palavras compostas* (ou *locuções*), cujo significado e construção são muitas vezes *não composicionais* (ou *idiomáticos*) – isto é, não decorrem das regras gerais de composição sintática e semântica que presidem às

combinações livres dos elementos lexicais. Nestas expressões não é possível, pois, determinar o significado global e as propriedades formais da construção a partir do significado e das propriedades dos seus elementos componentes, quando estes são usados separadamente na língua. Estas locuções abrangem igualmente, todas as categorias morfossintáticas, a um nível sintático *intrafrásico*. Falamos, assim, de nomes compostos (e.g. *golpe de estado*) e das locuções adjetivais (e.g. *mal-educado*), adverbiais (e.g. *sem dúvida*), conjuncionais (e.g. *posto que*) e preposicionais (e.g. *até a*); mas também expressões de natureza pronominal (e.g. *a gente*), determinativa (e.g. *a maioria de, um milhão de*) e interjetiva (e.g. *Cruz-credo!*)⁹. A fixidez combinatória entre estes elementos permite determinar que se trata de novas unidades lexicais, com estatuto idêntico ao das palavras simples, embora nem sempre o seu tratamento lexicográfico seja tão exaustivo como o que já se atingiu para as unidades lexicais simples. Já a um nível *frásico*, a composição pode igualmente abranger as construções verbais, quando o verbo e um ou mais dos constituintes da frase são fixos entre si, dando a combinação origem a uma nova construção verbal, muitas vezes surpreendente e idiomática: são as chamadas *locuções verbais*¹⁰ (ou *frases* ou *expressões fixas*, *expressões idiomáticas* ou *frases/expressões cristalizadas*) em que, muitas vezes, e tal como nas palavras compostas das categorias gramaticais elementares, não é possível derivar regularmente as propriedades distribucionais e transformacionais da construção das propriedades que o verbo apresenta quando usado noutras combinações. Tal como as chamadas palavras compostas, estas construções verbais fixas são mais difíceis de determinar, sendo necessário contrastar a fixidez formal (sintática) observada com a variação distribucionalmente livre que se verifica nas construções dos verbos simples por que são formadas. A idiomaticidade é, frequentemente, um bom índice dessa fixidez combinatória, mas a sua extensão e complexidade pode variar. Na grande maioria dos casos, apesar da fixidez da combinatória do verbo com um dos elementos da construção, um (ou mais) dos argumentos do verbo é distribucionalmente livre. Por exemplo, na maioria das expressões idiomáticas do português já estudadas (Galvão et al. 2019a, 2019b) o sujeito é distribucionalmente livre. No caso das construções com sujeito fixo, muito menos numerosas, não é raro haver um ou mais constituintes livres.

Quando comparamos a construção de *meter* na frase *O Pedro meteu o livro na pasta* (38LD)¹¹ com uma outra construção da mesma palavra, *O Pedro meteu-se com o João* (35R); as diferentes construções deste verbo, cada uma associada a um significado e construção próprias; cujas posições argumentais (sujeito e complementos) podem ser

⁹No caso das interjeições, a fronteira entre unidade léxica intrafrásica e unidade ao nível da frase elementar esbate-se consideravelmente, já que muitas interjeições formam frases completas.

¹⁰Este termo não se confunde com o que adiante usam Cunha & Cintra (1984), para as construções das chamadas perífrases verbais.

¹¹Os códigos alfanuméricos identificam a classe de construções verbais do léxico-gramática dos verbos do português europeu, descritas por Baptista (2012, 2013) e compiladas em Baptista & Mamede (2020).

consideradas como livres, isto é, com uma distribuição característica e determinável em compreensão; e cujas propriedades transformacionais dependem de cada uma dessas construções particulares; levam, naturalmente, a considerar que o conceito de unidade lexical pode ser aqui concebido como o de *unidade léxico-sintático-semântica*. De resto, esta é, em poucas palavras, a posição teórica do Léxico-Gramática (M. Gross 1975, 1982, 1988b), modelo de inspiração harrissiana (Harris 1964, 1965, 1982, 1991), para a qual a unidade lexical não é a palavra mas sim a frase elementar (ou a construção) de um elemento predicativo, como, neste caso, o verbo (simples) *meter*. Falamos, assim, para estas duas construções diferentes do verbo *meter*, de duas unidades lexicais distintas, ou, mais propriamente, de duas unidades *léxico-sintático-semânticas* diferentes.

Assim, é perfeitamente natural que se atribua às expressões fixas o mesmo estatuto de unidade lexical, à semelhança do que se faz para as construções livres do mesmo elemento principal. Por exemplo, em *O governo meteu o projeto na gaveta* (CNP2)¹², a combinação fixa *meter na gaveta* tem uma interpretação idiomática, não composicional, que autoriza, por exemplo, um coletivo humano (governo) como sujeito, o que não sucede na construção livre correspondente (*O Pedro meteu o livro na pasta*).

1.1.2 A unidade provérbio e outros tipos de unidade no sistema linguístico

Finalmente, neste quadro, ainda apenas muito sucintamente esboçado, podemos agora colocar a questão: *Como integrar os provérbios no sistema da língua?* À semelhança das expressões fixas, os provérbios também apresentam fixidez combinatória entre os seus elementos componentes, mas o que os distingue das expressões fixas é o facto de não apresentarem nenhuma posição sintática distribucionalmente livre. Ainda que possa haver variação (e esta variação pode assumir aspetos bastante complexos), ela limita-se a um pequeno paradigma de elementos, nem sempre distribucionalmente motivado, isto é, definível em compreensão. Por outro lado, um provérbio tanto pode ser idiomático, como, por exemplo, em:

Gato escaldado de água fria tem medo;

como pode ter um significado absolutamente literal, como em:

Faz o que eu digo e não faças o que eu faço.

¹²Os códigos alfanuméricos identificam a classe do léxico-gramática das construções verbais fixas do português europeu, descritas por Baptista et al. (2004) e Baptista et al. (2016), cuja última versão se pode encontrar em Galvão et al. (2019a, 2019b)

Para tal alerta Lopes (1992), relativamente a esta questão da idiomaticidade,

“que há provérbios idiomáticos, aqueles que desencadeiam uma interpretação-padrão não dependente dos princípios de uma semântica composicional, mas há também provérbios cuja interpretação-padrão coincide com o seu significado literal, composicionalmente apreendido” (*ibidem*: 23).

1.1.3 Uma visão tradicional

A descrição acima não se afasta muito de uma visão do léxico a partir da posição da gramática tradicional, entendida aqui como a gramática de tradição greco-latina, associada também à noção de *gramática normativa*, que rege o conjunto de regras associadas ao bem falar e ao bem escrever, segundo o uso e a autoridade de escritores e de gramáticos, entre outros, e de que são exemplo a gramática de Cunha & Cintra (1984) e de Bechara (2010).

Nesta perspetiva, os diferentes tipos de unidades lexicais fixas aparecem associados a diferentes denominações consoante a categoria gramatical a que pertencem. Cunha & Cintra (1984) consideram, assim, as *locuções conjuntivas* (p. 586), *locuções adverbiais* (p. 540) e *locuções prepositivas* (p. 552) as sequências coesas de elementos lexicais que apresentam na sua estrutura, respetivamente, conjunções, advérbios e preposições ligados a outros elementos gramaticais. Já as *locuções* (p. 393) (ou *perífrases*) *verbais* correspondem, para os autores, às expressões formadas de um verbo auxiliar com um verbo principal no participípio, no gerúndio ou no infinitivo impessoal, eventualmente ligados por uma preposição (e.g. *ter feito*, *estar fazendo*, *poder fazer*)¹³. A mesma gramática tem em atenção os nomes e os adjetivos que surgem ligados a outros elementos lexicais, denominando-os nomes compostos e adjetivos compostos. No entanto, esta gramática não chega a determinar um estatuto específico para as expressões idiomáticas nem, muito menos, para os provérbios, no quadro da definição das unidades do léxico da língua.

Já Bechara (2010: 493) faz corresponder expressão idiomática a idiotismo:

“idiotismo ou expressão idiomática é toda a maneira de dizer que, não podendo ser analisada ou estando em choque com os princípios gerais da Gramática, é aceita no falar formal”.

O autor dá como exemplos de idiotismos “a expressão *é que*, o infinitivo flexionado, a preposição *de* na expressão avaliativa/afetiva *o bom do pároco*, etc.” (*ibidem*). Para

¹³Uma gramática deste tipo de construções com auxiliares foi desenvolvida e implementada num sistema de processamento computacional do português por Baptista et al. 2010.

melhor explicar este conceito, refere Said Ali “Não devemos definir o idiotismo, segundo alguns gramáticos, como construção particular de uma língua, estranha, portanto, às outras línguas, porque ninguém conhece todos os outros idiomas em todos os seus segredos e modos especiais de falar”. (*apud* Bechara *ibidem*). Bechara conclui essa secção da gramática com a explicação de que o infinitivo flexionado é, pois, um idiotismo não porque só exista no português, já que, na realidade, existe em outras línguas (por exemplo, no húngaro), mas sim “porque a sua flexão contraria o conceito de forma infinitiva (isto é, não flexionada)” (*ibidem*).

Será interessante dizer que esta definição surge na parte da gramática que trata os *vícios e anomalias da linguagem*. Talvez o autor, ao falar de “choque com os princípios gerais da Gramática” pretendesse referir-se à idiomaticidade ou não composicionalidade semântica destas expressões. Não é esse o nosso entendimento, dado que os exemplos aduzidos invocam aspetos de natureza formal: (i) a forma expletiva *é que*, cujo funcionamento complexo até poderia ser aproximado, em muitas situações, do de um advérbio focalizador (Molinier & Lévrier 2000; veja-se também Costa 2008); (ii) a morfologia do infinitivo flexionado; e (iii) a chamada *construção cruzada* do adjetivo *bom*, que já foi sistematicamente descrita para os adjetivos intransitivos humanos do português europeu por Carvalho (2007), demonstrando a abundante extensão lexical do fenómeno.

1.1.4 O provérbio: Marginalidade e centralidade na língua

Esta visão da gramática pressupõe que as expressões idiomáticas seriam uma espécie de estruturas *marginais* da língua, no duplo sentido deste adjetivo: (i) no sentido em que não seguem os princípios gerais de boa formação sintática das frases e sintagmas da língua; e (ii) no sentido de serem em número residual, quando comparadas com as chamadas construções livres. Ora, sabemos hoje que tal não é assim. Não só a grande maioria das expressões idiomáticas e das palavras compostas de todas as categorias gramaticais seguem, de um modo geral, os princípios de boa formação das combinatórias da língua, como também o seu número é verdadeiramente expressivo face ao vocabulário das palavras simples. Quanto ao primeiro aspeto, veja-se o esquema de classificação das construções fixas desenvolvido no quadro do Léxico-Gramática (para o francês: Gross (1982); para o português do Brasil: Vale (2001); para o português europeu: Galvão et al. (2019), que segue de perto – com as devidas adaptações – as estruturas sintáticas das frases elementares (de núcleo verbal) da língua. Aliás, é esta conformidade das formas sintáticas das expressões idiomáticas de núcleo verbal (mais de 2000 expressões já recenseadas) às regras de boa formação sintática das frases elementares que permitiu desenvolver uma proposta de análise automática destas expressões (Rassi, Santos-Turati, et al. 2014) baseada numa análise prévia, genérica,

das frases, e implementada num sistema de processamento computacional da língua (Baptista et al. 2016, Galvão et al. 2019a, 2019b).

Também, pela nossa parte, foi possível, neste trabalho, propor uma classificação sintática dos provérbios (Reis & Baptista 2019b), que segue, na prática, a estrutura geral das frases da língua, revelando que a maioria dos provérbios por nós compilados não se distingue sintaticamente das formas de frase características da língua.

Por exemplo, M. Gross (1982) refere mais de 30 mil construções verbais idiomáticas, número que contrasta com os 12 mil empregos de verbos em construções distribucionalmente livres. O número de nomes compostos, na perspectiva de G. Gross (1988), especialmente se considerarmos as línguas de especialidade será tão grande ou até mais elevado do que o vocabulário das palavras simples da língua. Naturalmente, o recenseamento dessas locuções e expressões idiomáticas não está ainda tão desenvolvido nos trabalhos lexicográficos atuais que permita ter-se uma ideia mais precisa dos números em jogo, mas certamente não é residual.

Que tenhamos conhecimento, nenhuma destas gramáticas acima referidas apresenta as expressões proverbiais enquanto unidades lexicais da língua. Cunha & Cintra, de resto, não dedicam nenhuma secção da sua gramática nem a frases fixas, nem a provérbios. Bechara faz uma breve alusão aos provérbios no capítulo sobre a fonética e fonologia, quando trata a aliteração, fenómeno que pode ser considerado como um processo mnemónico, para ajudar a retenção destas expressões, dada a natureza predominantemente oral da sua transmissão cultural. O autor faz ainda referência aos provérbios no capítulo sobre os enunciados sem núcleo verbal, quando descreve as *frases assertivas bimembres* de que dá como exemplos os provérbios *Casa de ferreiro, espeto de pau* e *Tal pai, tal filho*. Contudo, ainda que o autor utilize estas expressões no texto como exemplos, não lhes chega a atribuir um estatuto linguístico específico, ou seja, não as reconhece como unidades lexicais, equiparáveis às palavras compostas ou às expressões idiomáticas.

1.1.5 Outra perspectiva

Ainda em relação ao tratamento de fenómenos de fixidez e as suas relações com o estatuto das expressões no léxico de uma língua, vejamos de seguida a perspectiva de outros autores sobre este tema.

Para Saussure (1916), parece-nos que a determinação do estatuto dos provérbios pode ser enquadrada na discussão que faz sobre a noção de *sintagma*. Para o autor, o conceito não se aplica apenas às palavras, mas também aos grupos de palavras, às unidades complexas de todas as dimensões e de toda a espécie, que ele especifica como palavras compostas “mots composés”, palavras derivadas “derivés”, membros da frase “membres de phrase” e frases inteiras “phrases entières” (*ibidem*: 172). Um

grande número de expressões pertencentes à língua são locuções fixas (“locutions toutes faites”), que o uso (“l’usage”) não autoriza que possam ser alteradas, ainda que nelas se possam distinguir, mediante reflexão, elementos constituintes portadores de significado (*ibidem*: 172). São exemplos do autor expressões interjetivas como “À Quoi, bon? Allons, donc!, etc.”, mas que são logo seguidas de exemplos de expressões fixas como as que acima referimos: “prendre la mouche” ‘ficar subitamente zangado com algo de pouca importância’ (cp. ‘fazer uma tempestade num copo de água’), “forcer la main à quelqu’un” ‘obrigar alguém a agir’.

No entanto, o autor refere (*ibidem*: 173) que é difícil classificar uma combinação de unidades, pelo facto de serem produto da *langue*, marcada pelo uso coletivo, e da *parole* que depende da liberdade individual. Este tipo de expressões (“tours”) apresentam características particulares de significado ou de construção, não podendo ser “improvisadas”, pois são produto de uma tradição (“ils sont fournis par la tradition”). Saussure (*ibidem*: 172-173) refere, ainda, que há palavras que são caracterizadas por alguma anomalia morfológica mantida apenas pela força do uso.

Não é, pois, completamente evidente se, para o autor, as expressões fixas deveriam fazer parte da *langue* (e, logo, do léxico) ou se são um mero produto da *parole*, resultado da variação produzida pela liberdade na expressão individual do falante. Parece-nos, contudo, que, pelas referências à impossibilidade de serem “improvisadas”, bem como pelo facto de serem “produto de uma tradição”, o estatuto das expressões fixas no sistema linguístico as aproxima mais de um elemento da *langue*, e assim do léxico. Contudo, e de um modo inconclusivo, Saussure parece indicar não existir, nestes casos, uma fronteira clara entre *faits de langue* e *faits de parole* (p. 173):

“[...] il faut reconnaître que dans le domaine du syntagme il n’y a pas de limite tranchée entre le fait de langue, marque de l’usage collectif, et le fait de parole, qui dépend de la liberté individuelle. Dans une foule de cas, il est difficile de classer une combinaison d’unités, parce que l’un et l’autre facteurs ont concouru à la produire, et dans les proportions qu’il est impossible de déterminer”.

1.1.6 Provérbios e tradução

Jakobson (1987) afirma que “[i]n any language there exist also coded word groups called phrase words” (*ibidem*: 98). Segundo o autor, “the meaning of the idiom cannot be derived by adding together the meanings of its lexical constituents” (*ibidem*), por outras palavras, o significado dessas expressões da língua não pode ser derivado da soma do significado de cada um dos seus constituintes lexicais. O autor refere, também (*ibidem*), que esses “grupos de palavras” se comportam como as unidades simples da língua e que, para compreender a sua maioria, é apenas necessário estar-se

familiarizados com as palavras que os constituem e com as regras sintáticas das suas combinações. Se se tiver em conta essas limitações, tem-se a liberdade de colocar as palavras em novos contextos, mas acrescenta que essa liberdade é relativa, já que a pressão dos clichés sobre a nossa escolha de combinações é considerável.

Ainda no que respeita às combinações de unidades linguísticas, o autor (*idem: ibidem*) considera que estas apresentam uma escala ascendente de liberdade, sendo a liberdade do falante nula no que se refere à combinação de traços distintivos em fonemas; na combinação de fonemas em palavras essa liberdade é limitada à invenção de novas palavras no léxico; na formação de frases com palavras, o falante é menos constrangido; e, por último,

“in the combination of sentences into utterances, the action of compulsory syntactical rules ceases, and the freedom of any individual speaker to create novel contexts increases substantially, although again the numerous stereotyped utterances are not to be overlooked.”

Não é evidente que se possa considerar os provérbios como estando abrangidos pelo termo “stereotyped utterances” (expressões estereotipadas), mas, certamente, parece-nos razoável admitir que as expressões fixas, tal como aqui as definimos, devem fazer parte deste tipo de expressões.

Num outro ponto do seu estudo (*ibidem: 429*), Jakobson volta a referir as *phrase words* a propósito do facto de estas unidades complexas das línguas poderem levantar diversos problemas no que se refere à sua tradução.

No quadro de uma abordagem semiótica da definição da forma como as expressões das línguas significam, o autor postula três tipos de tradução: (i) intralinguística (“rewording” ou paráfrase); (ii) interlinguística (tradução propriamente dita); (iii) e intersemiótica, quando se trata de passar de um sistema semiótico de signos para outro. Focando-se no conceito de tradução como paráfrase, o autor afirma que:

“A word or an idiomatic phrase word, briefly a code unit of the highest level, may be fully interpreted only by means of an equivalent combination of code units”,

ou seja, identifica “a word or a idiomatic phrase word” (expressão idiomática) como “a code unit of the highest level” (uma unidade do código, do mais alto nível), pelo que só pode ser interpretada por uma combinação equivalente de “code units” (unidades do código). Já ao nível da tradução interlinguística, “there is ordinarily no full equivalence between code units” (*ibidem*), mas sim a transferência de uma língua para a outra de interpretações de unidades ou das mensagens:

“Most frequently, however, translation from one language into another substitutes messages in one language not for separate code units but for entire messages in some other language” (*ibidem*: 430).

Parece, pois, depreender-se que, as expressões fixas, que aqui identificamos com o termo jakobsiano “idiomatic phrase word”, constituem assim unidades do sistema linguístico (código), pelo que seria natural enquadrar do mesmo modo os provérbios enquanto mensagens que deverão ser tomadas como um todo, especialmente tendo em conta as dificuldades que a sua tradução, tanto intra- como interlinguística levantam.

1.1.7 Uma visão integradora da dimensão discursiva dos provérbios: Zuluaga (1980)

Um autor incontornável nesta problemática do estatuto dos provérbios no sistema linguístico é, provavelmente, Alberto Zuluaga, quer pela minuciosa revisão que faz dos trabalhos anteriores no quadro da gramática generativa-transformacional (Zuluaga 1975a, 1975b), quer também pela forma como define e enquadra as unidades fraseológicas numa teoria gramatical cogente (Zuluaga 1980).

Segundo o autor, as unidades fraseológicas podem ser agrupadas em dois grandes grupos: *locuções* e os *enunciados* [*fraseológicos*]. Das locuções fazem parte os *instrumentos gramaticais*, as *unidades léxicas* e os *sintagmas*, todos estes com as respetivas subdivisões (cf. Zuluaga 1980: 139); dos enunciados fazem parte, por um lado, as *frases* (que incluem os *clichés*, as *fórmulas* e os *ditos*) e os *textos* (*provérbios*, a que chama “refranes”). Enquanto os enunciados “no requieren un contexto verbal inmediato para constituir una expresión de sentido completo en el habla”, as locuções “requieren contexto verbal inmediato, es decir, se combinan con otros elementos en el interior de la frase” (itálico nosso). Esta definição vai ao encontro, pois, da que já vimos atrás e em que distinguimos as chamadas *palavras compostas* (nome, adjetivo, etc.), que são elementos de formação da frase, das *expressões fixas*, as construções que determinam uma frase e são a expressão de um predicado semântico completo.

Zuluaga considera, depois (*ibidem*: 191), que os enunciados fraseológicos são expressões fixas equivalentes ou superiores à frase e que tanto podem corresponder a uma oração (simples ou complexa), como podem consistir em um sintagma ou em uma simples palavra (usada com um valor equivalente ao de uma frase): o seu traço definatório é funcionarem como *unidades comunicativas mínimas*, com sentido próprio, enunciadas pelo falante entre duas pausas e em unidades de entoação distintas.

De acordo com a sua estrutura interna, o autor (*ibidem*: 192-198) classifica os enunciados fraseológicos de acordo com três parâmetros: (i) idiomaticidade, (ii) nível de estruturação gramatical, e (iii) modalidade. No que respeita à *idiomaticidade*, o autor subclassifica os enunciados fraseológicos em:

- (a) *Meramente fixos ou de sentido literal*, como por exemplo, “dime con quien andas y te diré quien eres” (*idem*) (cf. *Diz-me com quem andas e dir-te-ei quem és*);
- (b) *Semi-idiomáticos*, isto é, os que apresentam um sentido simultaneamente literal e idiomático. Neste caso, o primeiro (sentido literal) serve de motivação funcional-sincrónica e não apenas etimológica ao segundo (sentido idiomático), como, por exemplo, “ojos que no ven, corazón que no siente (= lo que no llega a nuestro conocimiento no tiene por qué afectarnos)” (*idem*) (cf. *Olhos que não veem, coração que não sente*); e
- (c) *Idiomáticos*, ou seja, os que não apresentam, funcionalmente, motivação linguística, como por exemplo, “contigo, pan y cebolla (= me casaría contigo, aunque tuviese que vivir en estrechez material)” (*idem*) (cf. *Amor e uma cabana*).

Relativamente ao segundo parâmetro, o nível de estruturação gramatical dos seus componentes, esses enunciados (a que o autor chama agora fixos) podem ser constituídos por:

- (1) Uma só palavra, como, por exemplo, “salud”;
- (2) Um sintagma nominal, como, por exemplo, “muchas gracias”;
- (3) Uma oração simples, nominal ou verbal, como, por exemplo, “a mula regalada no se le mira el diente” (cf. *A cavalo dado não se olha o dente*);
- (4) Uma oração composta, tanto por *hipotaxe*, como, por exemplo, “no hagas a otros lo que no quieras que hagan contigo”, (cf. *Não faças aos outros [o que não queres que te façam a ti]*), como por *parataxe*, como, por exemplo, “haz bien y no mires a quien”; (cf. *Faz o bem sem olhar a quem*);
- (5) Uma citação introduzida por uma oração com um ‘verbo de expressão linguística’ (este tipo de enunciados fixos também são designados por *Wellerismos*¹⁴, como, por exemplo, “dijo la zorra a las uvas: no están maduras” (cp. *Diz o roto ao nu: Porque não te vestes tu?*);
- (6) Uma oração interrogativa unida fraseologicamente a uma resposta. Estes enunciados são também designados por *dialogismos*¹⁵, como, por exemplo, “Quién te hace rico? El que te mantiene el pico” (cp. *Quem te fez rico? O pão da minha aldeia*; Parente LP_SP35425)¹⁶;

¹⁴Segundo Zuluaga (1980: 253) “Del personaje Sam Weller de Dickens”.

¹⁵cf. Zuluaga (1980: 195).

¹⁶O código convencional aqui indicado identifica o provérbio na base de dados de provérbios, estabelecida a partir de quatro coletâneas de textos contemporâneas, que será apresentada mais adiante neste trabalho.

Segundo o autor (*ibidem*: 195), os últimos quatro tipos de enunciados apresentados (3-6) são normalmente conhecidos por provérbios (“refranes”) e são definidos como verdades tradicionais de validade geral e permanente.

O autor (*ibidem*) chama ainda a atenção para os enunciados que tem vindo a designar por orações simples nominais, como “buen cobrador, mal pagador” (cf. *Todo o bom cobrador é mau pagador*), “cada loco con su tema” (cf. *Cada maluco com sua mania*) ou “dicho y hecho” (cf. *Dito e feito*), e refere que estes são, por norma, descuidados em estudos de provérbios. Zuluaga considera-os provérbios prototípicos, embora defenda que os mesmos devem ser entendidos como orações completas, uma vez que contêm uma mensagem estruturada de forma idêntica à das orações com predicado verbal.

Finalmente, o autor estabelece o parâmetro da *modalidade*, que diz respeito ao tipo de relação social estabelecida mediante a produção de um enunciado. Essa relação pode ter diferentes valores gerais: declarativo (informação), interrogativo (pergunta) ou imperativo (exortação). Zuluaga (*ibidem*: 198) apresenta, efetivamente, enunciados fraseológicos sob a forma de orações declarativas, e.g. “um clavo saca otro clavo” (cf. *Um prego empurra outro*); JP Machado GLP_JPM27204), imperativas (p. ex. “haz bien y no mires a quien” (cf. *Faz o bem sem olhares a quem*) e interrogativas, e.g. “Que sabe el burro de la miel?” (cf. *Quem te manda a ti sapateiro tocar rabecão?* JP Machado GLP_JPM24027). Note-se, porém, que estes últimos enunciados, interrogativos, são apenas questões de caráter retórico, funcionando geralmente como declarações ou comentários. O autor (*ibidem*) considera, aliás, que estas modalidades poderão ser “muchísimo más variadas” e que só podem ser efetivamente reconhecidas se tivermos em conta o contexto (entendido enquanto contexto situacional/comunicativo, ou seja de ordem pragmática) em que o enunciado fraseológico é proferido.

Em relação ao comportamento funcional dos enunciados fraseológicos, Zuluaga (*ibidem*: 199-200) verifica que estes enunciados podem ser diferentes entre si, podendo ser separados em grupos ou classes consoante o contexto em que se podem integrar e segundo as relações que podem estabelecer com esses mesmos contextos. Neste sentido, o autor distingue os enunciados fraseológicos *funcionalmente livres* dos enunciados fraseológicos *contextualmente marcados*. Os primeiros caracterizam-se por não dependerem de nenhum tipo de contexto linguístico nem extralinguístico para serem plenamente entendidos; enquanto que os segundos dependem de algum contexto ou de algum fator da fala para adquirirem um sentido completo. Zuluaga (*ibidem*: 201) inclui os provérbios nesse primeiro grupo por conterem uma mensagem ou informação com sentido completo, daí os considerar *textos mínimos*.

Vemos, assim, que Zuluaga esbate, ao descrever os provérbios, as distinções entre o conceito de unidade lexical e o de enunciado: o provérbio será uma unidade da língua, mas que funciona de um modo particular como um texto dentro de outro texto maior,

e que com ele estabelece complexas relações de sentido, em parte capturadas pela noção (pragmática) de *modalidade* a que o autor recorre. O autor usa uma abordagem predominantemente formal para a sua classificação das unidades fraseológicas. Os parâmetros usados para descrever (e classificar) as diferentes estruturas encontradas não são mutuamente exclusivos: (i) a *idiomaticidade* não é um critério absoluto, que permita distinguir este tipo de unidades; (ii) o *nível de estruturação gramatical* abrange não só expressões formadas por uma só palavra, que constitui um enunciado completo, até expressões mais longas de ordem frásica, que podem inclusive apresentar processos de hipotaxe e parataxe; e, até, expressões que simulam pequenos diálogos, como no caso dos chamados *wellerismos*, que envolvem um verbo *dicendi*, (Baptista 2010), ou nos chamados *dialogismos*; e, finalmente, (iii), a introdução do conceito de *modalidade* – entendido como o valor pragmático do enunciado – que conceptualiza a expressão proverbial e outros enunciados fraseológicos não apenas como combinações de elementos lexicais fixos entre si, mas também expressões em que a própria função pragmática se cristaliza com o material morfémico do enunciado.

A ideia de que a modalidade é um dos aspetos inerentes à propriedades das unidades lexicais não é inteiramente nova, como o próprio Zuluaga reconhece (*ibidem*: 198), citando Benveniste (1962), dizendo que esta categoria “se basa en la predicación”. Efetivamente, ainda que para a maioria das construções, simples e compostas, da língua, a modalidade neutra (declarativa) possa ser considerada a forma de base, verifica-se que há numerosas construções, sobretudo fixas e idiomáticas, em que a modalidade está cristalizada com a própria expressão. Gross (1982) refere, por exemplo, toda uma classe de construções “exclamativas”, que não podem ser usadas sem esse valor: e.g. *Minute, papillon!* Esta noção pode ser posta em relação com o conceito de *pragmatema* (Mel’čuk 1995), que encontramos, entre outros, em Iriarte Sanromán (2001), Blanco (2010) e Pinheiro (2015). Determinar, contudo, os valores pragmáticos precisos que é possível associar aos provérbios não é uma tarefa fácil, como procurámos demonstrar em Reis & Baptista (2019a).

1.1.8 Contribuição do modelo do Léxico-Gramática

A descrição linguística dos provérbios encontra um quadro natural no modelo teórico-metodológico do Léxico-Gramática (M. Gross 1975, 1982, 1988b), desenvolvido no quadro da gramática transformacional de operadores de Z. S. Harris (Harris 1964, 1965, 1982, 1991). Não nos alongaremos aqui na descrição deste modelo; veja-se, por exemplo, Lamiroy (1998); até porque já nos referimos à forma como são conceptualizadas as expressões fixas neste enquadramento; veja-se, por exemplo, as reflexões de Ranchhod (2003) a este respeito e com referência ao português. O ênfase posto pelos investigadores ligados a esta perspetiva na descrição deste tipo de

expressões é notável, tendo-se produzido extensos léxicos de palavras compostas de várias categorias gramaticais: nomes compostos (G. Gross 1988), advérbios (M. Gross 1996); e, sobretudo, extensos léxico-gramáticas de frases fixas (1982). Para o português, é possível indicar diversos estudos já desenvolvidos, neste quadro teórico, sobre as expressões idiomáticas de diferentes tipos (nomes compostos: Baptista (1994); expressões idiomáticas verbais: Baptista et al. (2004, 2005, 2016), Fernandes (2007), Galvão et al. (2019a, 2019b); advérbios compostos: Ranchhod (1991); advérbios, preposições e conjunções compostas: Palma (2009), entre outros. Em relação aos provérbios, refira-se, por exemplo, Chacoto (1994, 2007, 2008 e 2010).

Tal como o define M. Gross (1994) “Lexicon-grammar is a model of syntax limited to the elementary sentences of a natural language”. No caso das frases livres, por exemplo, as construções verbais, estas são frases elementares, concebidas como “elementary units for syntactic as well as for semantic composition” (*ibidem*). Neste sentido, as frases fixas de núcleo verbal são expressões idiomáticas cuja descrição linguística não difere, no essencial, da descrição das construções verbais livres que envolvem esses mesmos verbos – apenas é necessário levar em conta o facto de uma (ou mais) posições estruturais ser distribucionalmente fixa com o verbo, pelo que se exige que seja representado explicitamente no léxico-gramática. M. Gross (1988a: 21, 22) relata que o estudo exaustivo de mais 30.000 frases fixas e sua comparação com um léxico de quase 12.000 frases livres, apoiado no léxico-gramática, lhe permitiu chegar às seguintes conclusões: (i) as frases fixas só excepcionalmente são completamente fixas; (ii) são quase todas analisadas de forma sintática regular; (iii) as regras sintáticas por que se regem são exactamente as mesmas regras das frases livres; e, por último, (iv) as frases fixas são mais numerosas do que as frases livres. O autor (1982: 161) refere, também, que há frases que não apresentam nenhuma posição livre, sendo inteiramente fixas. Nomeadamente, frases como *Les carottes sont cuites* e provérbios como *Tous les chemins mènent à Rome*. Maurice Gross (*ibidem*: 161-162) considera que, apesar destas duas frases serem totalmente fixas e não permitirem comutações, a intuição é clara no que concerne a distinguir as expressões fixas dos provérbios e que a intuição linguística do que é um provérbio (adágio, máxima ou sentença) parece estar ligada ao carácter geral da frase, conferido pelo uso de determinantes genéricos.

1.1.9 Provérbios e Genericidade

Efetivamente, o carácter genérico das expressões proverbiais resulta do emprego de diferentes processos gramaticais, não apenas ao nível dos determinantes, nomeadamente, o emprego dos chamados quantificadores universais (e.g. *todo*: **Todos os caminhos vão dar a Roma**; *cada*: **Cada macaco no seu galho**) ou dos artigos definidos sem valor referencial (*Pela boca morre o peixe*), mas também no emprego

de certos tempos-modos verbais, e.g. infinitivo impessoal (*Partir é morrer um pouco*) ou expressões com o verbo na 3ª pessoa do singular construído com pronome indefinido *se* (*Devagar se vai ao longe*). Estes tempos-modos e estas construções sintáticas contribuem para o caráter intemporal e sentencioso, habitualmente associado às expressões proverbiais. Justamente, um destes mecanismos é o emprego de pronomes indefinidos (e.g. *alguém, ninguém, qualquer*), de que tratamos mais aprofundadamente em (Reis & Baptista 2018 [corresponde ao Artigo 7 desta tese]). Contudo, a determinação do estatuto proverbial destas expressões pode envolver outros aspetos formais, tais como a estrutura bimembre (*Cá se faz, cá se paga*), o paralelismo interno (*Filho és, pai serás*), a aliteração (*Quem com ferro fere com ferro será ferido*), a rima (*Abril, águas mil*), etc. Esses processos podem ser, inclusive utilizados como índices para a reutilização dos provérbios em numerosas atividades didáticas, aspeto que tratamos especificamente em (Reis & Baptista 2016a [corresponde ao Artigo 2 desta tese]).

1.1.10 Estrutura das frases proverbiais

Ora, embora muitos provérbios apresentem uma estrutura sintática idêntica à de uma frase simples, com um único verbo (e.g. *Pela boca morre o peixe*); muitos há também que não apresentam qualquer núcleo predicativo (e.g. *Pão-pão, queijo-queijo*), enquanto outras podem ser associadas a frases complexas com mais do que uma proposição/oração (e.g. *Cá se faz, cá se pagam*), incluindo processos de parataxe (e.g. *Faz o que eu digo e não faças o que eu faço*) e hipotaxe (e.g. *Quando a esmola é muita, o pobre desconfia*). Esta diversidade formal, sintática foi objeto de descrição sistemática em (Reis & Baptista 2019b [corresponde ao Artigo 8 desta tese]). Esta abordagem taxonómica permite circunscrever melhor certos fenómenos sintáticos específicos, frequentemente numa perspetiva contrastiva ou de tradução (por exemplo, as construções interrogativas indiretas introduzidas por *quem* (Conenna 1988, Chacoto 2007); ou as construções comparativas (Chacoto 2008), ou as estruturas condicionais com *se* (Chacoto 2010); e descrever, de forma não casuística mas sim transversal, conjuntos formalmente mais homogêneos de construções, relevando propriedades comuns e aspetos da sua variação.

Efetivamente, os provérbios que não apresentam uma estrutura frásica mais regular levantam dificuldades a modelos de análise gramatical mais tradicional, que pretendem ver neles formas reduzidas de estruturas mais longas, completas, de onde se poderia derivar o significado global da expressão. Tome-se como exemplo Bechara (2010: 412), para quem a oração

“constitui o centro da atenção da gramática por se tratar de uma unidade onde se relacionam sintaticamente seus termos constituintes e onde se

manifestam as relações de ordem e regência, que partem do núcleo verbal, e das quais se ocupa a descrição gramatical.”

Por isso,

“[s]e um enunciado se expressa sob forma de oração, a gramática tem condições de surpreender nela as relações que existem entre si. A apreensão, descrição e sistematização dessas relações constituem o objeto da investigação da gramática descritiva” (*ibidem*: 4).

Ora, de acordo com Bechara (2010: 2), “a gramática pouco tem de dizer diante de enunciados como: *Bom dia! Adeus.*”, enquanto que em enunciados como “*Tal pai, tal filho*” a gramática pode ter em consideração a classe gramatical, o gênero e o número dos seus constituintes e estabelecer possibilidades de equivalência com enunciados como, por exemplo, “*Tal é o pai, tal é o filho*” ou “*Tal se parece o pai, tal é (ou se parece) o filho*”.

Ainda em relação aos enunciados sem núcleo verbal, Bechara refere que frases como *Casa de ferreiro, espeto de pau* “diferem da oração porque são proferidas, quase sempre, em situações especiais, fora das quais a comunicação não se concretiza em toda a sua plenitude” (*ibidem*: 412). Na impossibilidade de analisar este tipo de expressões proverbiais, ou seja, de determinar as relações sintáticas entre um (eventual) verbo e os seus sujeito e complementos, o autor propõe que sejam apenas descritos “de acordo com a sua constituição interna”, por exemplo, pelo número de membros (às vezes chamados *hemistíquios*) por que são formados.

Apesar desta diversidade formal, a fixidez combinatória que caracteriza cada provérbio leva a que estas expressões tenham de ser aprendidas *como um todo*, incluindo as condições pragmáticas que presidem à sua utilização em discurso (ou em contexto comunicativo). Por essa razão, parece-nos mais adequado atribuir-lhes o estatuto de unidade, de valor idêntico ao das unidades léxico-sintáticas, mas mais complexa, por envolver também, e de forma sistemática, uma componente pragmática.

1.1.11 Provérbios e unidades paremiológicas (UP)

Apesar da fixidez combinatória que exibem, a variação formal que os provérbios admitem é provavelmente um dos maiores desafios que estas estruturas levantam quer ao seu tratamento lexicográfico, quer, numa perspetiva mais ampla, ao seu reconhecimento automático em textos.

Um dos tipos de variação mais frequentes é a chamada *variação lexical*, quando um dos elementos do provérbio pode ser substituído por outra palavra sem que o significado global (e as condições pragmáticas de uso, definidas de forma genérica) se alterem. Esta

variação lexical por vezes acarreta ligeiros ajustamentos morfossintáticos para garantir a concordância entre os membros da frase, e.g.

A cavalo/burro dado não se olha/olham o dente/os dentes.

Neste caso, a variação *cavalo/burro* não tem efeito sobre o resto da frase, mas a alternância *dente/dentes* conduz à variação na flexão de *olhar* e do determinante. Outras formas de variação implicam restrições combinatórias a maior distância:

A vaca/galinha da vizinha é mais gorda do que a minha.

A vaca da vizinha dá mais leite do que a minha.

Naturalmente, se se falar do par *vaca-leite* ou do par *vaca/galinha-gorda*, as condições contextuais para a escolha de uma ou outra variante poderão ser diferentes.

Em certos casos, um elemento pode ser facultativamente elidido sem que o provérbio sofra uma alteração do significado; e.g. a palavra *pássaro* no provérbio seguinte

Mais vale um (pássaro) na mão do que dois no ar/a voar.

Finalmente, certas formas de variação podem modificar a forma da frase de modo sintaticamente mais profundo. Por exemplo, veja-se o que sucede com a chamada extração *ser que* (ou *clivagem*):

É a intenção que conta.

A intenção é que conta.

Não é só a intenção que conta.

que, neste provérbio, parece ser um processo obrigatório: **A intenção conta.*

De um modo semelhante, nas estruturas comparativas, diferentes formas de expressar a comparação produzem estruturas sintáticas diferentes:

Mais vale ser que parecer

Antes ser que parecer

Outro aspeto que torna muito complexa a descrição da variação formal dos provérbios é a possibilidade de membros inteiros poderem ser omitidos:

Uma mão lava a outra (e as duas lavam a cara)

Mãos frias, coração quente (amor ardente)

Neste último caso, o provérbio pode ainda conter um segundo verso, que estabelece vários paralelismos com o primeiro:

Mãos frias, coração quente, amor ardente;

mãos quentes, coração frio, amor vadio

Os autores que se dedicam à recolha de provérbios tentam, de alguma forma, mitigar o impacto deste fenómeno recorrendo a diferentes processos de notação, que tentámos emular nos exemplos acima, como por exemplo, o uso de barras oblíquas, parênteses, etc. Contudo, subjacente a esta variação encontra-se a ideia de que todas estas formas correspondem a diferentes expressões de uma mesma unidade de sentido, com idênticas condições pragmáticas de uso. A esse conceito abstrato, unificador das múltiplas variantes por que o mesmo conteúdo pode ser expresso chamamos *unidades*

paremiológicas (UP). Serão essas UP, e não as variantes individuais que uma mesma UP possa apresentar, que deverão figurar no léxico do provérbios. Como qualquer outra parte do léxico, o falante individual poderá conhecer apenas uma ou algumas dessas variantes, mas a comunidade linguística, como um todo, poderá reconhecer, mesmo em variantes menos usuais, essa identidade formal e essa invariância de significado e de condições de uso a que chamamos UP.

Ao longo deste estudo, procurámos aferir os critérios formais que nos permitiram constituir o léxico das UP do português europeu. Por não ter cabimento nesta introdução, apresentamos em anexo (Anexo 2) o conjunto de diretivas que elaborámos para, de forma sistemática e consistente, reunir sob uma mesma UP provérbios às vezes formalmente bem distintos.

1.1.12 Variação formal e classificação sintática

A variação interna dos provérbios é, como dissemos, uma dificuldade ao seu reconhecimento automático em textos. Foi para vencer esse desafio que, entre outros, Rassi, Baptista, & Vale (2014) constituíram, num primeiro momento, uma classificação (sintático-semântica) das diferentes estruturas observadas. Para tal, os autores usaram uma coleção de cerca de 3.500 provérbios e variantes, recolhida a partir de coletâneas de provérbios do português do Brasil. Depois, estudaram diferentes métodos alternativos para capturar diretamente em textos variantes previamente não registadas de provérbios, tanto em *corpora* como em listas recolhidas a partir de coletâneas especializadas.

Pela nossa parte, retomámos esta questão, agora sobre os provérbios do português europeu. Apoiados numa base de dados de 114 mil provérbios, reunidos a partir de quatro coletâneas de textos contemporâneas (Costa 2004, Machado 2011, Moreira 1996, Parente 2005), elaborámos uma nova classificação (Reis & Baptista [corresponde ao Artigo 8 desta tese]), estritamente formal (sintática) das estruturas encontradas nas expressões proverbiais, procurando constituir conjuntos o mais homogêneos que fosse possível do ponto de vista formal e visando um procedimento de classificação automática aplicável à generalidade da lista constante da base de dados. Na mesma linha, estudámos diferentes métodos de capturar os provérbios em textos, incluindo variantes não registadas, seja pelo recurso a palavras-chave, seja por uma flexibilização das restrições morfossintáticas impostas à representação das variantes dos provérbios. Desta dimensão do trabalho damos conta, por exemplo, em Reis & Baptista (2016b) e Reis & Baptista (2017b) [correspondem aos Artigos 3 e 4 desta tese].

1.1.13 Representação formal das unidades paremiológicas (UP)

Evidentemente, dada complexidade e a diversidade dos fenómenos de variação observados no seio de muitas UPs, a representação sob a forma de listagem, tradicionalmente adotada nas coletâneas é bastante inadequada, havendo outros processos de representação muito mais eficientes e precisos, tais como o emprego de grafos de estados finitos ou expressões regulares (que são outro formalismo, mas equivalente). Acresce que estes métodos podem ser diretamente explorados por sistemas de processamento de texto em língua natural e, assim, servir para identificar os provérbios nos textos em que ocorrem. Mais, estes métodos permitem a identificação não só dos provérbios mas também de todas as variantes que estes admitem (desde que tenham sido adequadamente formalizadas), relacionando essas diferentes estruturas e formas equivalentes sob uma mesma UP.

O reverso desta abordagem é, naturalmente, a necessidade de formalizar, manualmente, estes processos de variação. Trata-se de um trabalho moroso, que não pode ser automatizado (tem de ser feito por humanos) e, como qualquer tarefa deste tipo, que está sujeito a erros e a imprecisões. Impunha-se, pois, definir prioridades, já que, a partir das dimensões da nossa base de dados, a tarefa se mostrava demasiado extensa e, para certas expressões registadas nas coletâneas que lhe serviram de base, perfeitamente inútil, dada a raridade ou mesmo o arcaísmo dos provérbios ali recolhidos. Foi nesse sentido que elaborámos um mínimo paremiológico do português europeu, assunto de que tratamos a seguir. Por outro lado, esta seleção deverá poder ser usada para outras finalidades, de que damos conta, mais adiante.

1.1.14 O mínimo paremiológico (MP)

Quando se procura representar o léxico de uma língua, sobretudo tendo em vista uma aplicação computacional dessa representação, há, certamente, a necessidade de proceder a uma seleção do material a incluir/descartar, já que o uso de algumas expressões pode ser tão marginal ou raro que não justifique a sua inclusão num léxico computacional (M. Gross 1988b). Acresce que, para aferir com rigor a variação de um provérbio e, em contraponto, a sua pertença à mesma unidade paremiológica, o investigador deve conhecer suficientemente bem não só as diferentes formas do provérbio como as condições pragmáticas de uso que autorizam essa variação.

Socorremo-nos, para tal, do conceito de *mínimo paremiológico* (MP) de uma língua (Ďurčo 2015, Ďurčo & Meterc 2014, Meterc 2012, 2014, Mieder 1992, 1994, Permjakov 1973), entendido como o conjunto de provérbios que deverá não só conter os provérbios que são parte de um fundo comum dessa língua, como também aqueles que são efetivamente utilizados pelos falantes. O MP deve ainda ser considerado um objeto

dinâmico, uma lista que deverá ser atualizada ao longo do tempo, de modo a representar o conhecimento linguístico de uma dada comunidade, num dado momento sincrónico da evolução da língua.

A dificuldade em estabelecer esse conjunto de expressões prende-se, sobretudo, na interseção e interação entre diferentes critérios, já que não seria adequado considerar apenas um tipo de dados de frequência (*corpora* jornalístico vs. internet) ou de disponibilidade lexical, entendida como reflexo de uma introspeção, elicitável por meio de questionários a uma amostra representativa da população que forma a comunidade linguística-alvo (Reis & Baptista 2017a). Por outro lado, até no quadro de uma mesma UP, diferentes variantes poderão ter um estatuto diferenciado relativamente ao mínimo paremiológico, com algumas variantes a serem muito conhecidas e efetivamente usadas pela maioria dos falantes de uma comunidade linguística, enquanto outras serão de uso mais raro (ou mesmo desusadas). O trabalho apresentado em Reis & Baptista (2020) pretende, justamente, ser uma contribuição para o estabelecimento de um mínimo paremiológico do português europeu, contendo cerca de 320 unidades paremiológicas, sob as quais se identificam as respetivas variantes, indicando as mais usuais.

1.1.15 Utilização do MP para aplicações didáticas

Este novo instrumento poderá ter diversas aplicações. Uma delas tem a ver com o ensino da cultura e da língua portuguesas, tanto língua materna como língua estrangeira. Procedemos a um levantamento crítico da forma como as expressões proverbiais são abordadas quer nos atuais documentos ordenadores do ensino do português (Buescu et al. 2015, Grosso (Org.) 2011, Leiria (Coord.) 2008) e verificámos em *corpora* de manuais escolares diversificados como é que, em concreto, os autores desses manuais abordavam ou usavam (ou propunham que se usasse) os provérbios enquanto material linguístico, culturalmente relevante. Naturalmente, a utilização do mínimo paremiológico poderá, em princípio, contribuir para uma mais eficiente e consistente organização da forma como os *curricula* integram este tipo de expressões nos manuais escolares, sem prejuízo do recurso a outras ferramentas ou fontes. Foi neste sentido que, pela nossa parte, desenvolvemos um conjunto de jogos didáticos (Reis & Baptista 2016a) com base em provérbios, que permitem explorar esse riquíssimo material linguístico para a aprendizagem da língua.

1.1.16 Utilização do MP para aplicações terapêuticas/diagnóstico

Os atuais instrumentos de diagnóstico de certas afasias da linguagem usam muitas vezes provérbios, já que são estruturas da memória que são afetadas diferencialmente,

mesmo quando a capacidade de falar é diminuída (Cazelato 2003, 2008, Santos 2009, Santos et al. 2009, Sé 2011). Ainda que os provérbios sejam utilizados para esse propósito de diagnóstico e até de terapia, nem sempre são descritos os procedimentos que conduziram à seleção apresentada nesses instrumentos (cf. Artigo 6, Reis et al. 2017). Considerando que os provérbios servem, muitas vezes, o objetivo terapêutico de eliciação de palavras e treino da memória, será pois determinante que essa escolha corresponda exatamente aos provérbios que são mais usuais na língua. Isto poderá resolver, em parte, a questão de o não reconhecimento de um determinado provérbio por um paciente se poder dever, não à sua patologia, mas sim ao seu desconhecimento do provérbio que lhe é apresentado.

A determinação de uma lista de provérbios muito usuais (v. Anexo 6) poderá, assim, servir como ponto de partida para o desenvolvimento de um conjunto de tarefas que sirvam esse objetivo diagnóstico/terapêutico específico. Essas tarefas poderão consistir, por exemplo, em pedir ao paciente que complete um provérbio de que se apresenta apenas uma das partes, ou mesmo que complete o provérbios quando se fornece uma palavra-chave previamente omitida (v. Anexo 7 e Anexo 8).

Neste sentido, este trabalho poderá ser um contributo não só por ser disponibilizada uma lista de provérbios que poderá servir de referência para os especialistas (ver Reis et al. 2017, [Artigo 6], Reis & Baptista [Artigo 9]), como também por apresentar algumas tipologias de jogos, cujo desenvolvimento assentou em processos linguísticos motivados e que refletem os fenómenos da língua (cf. Artigo 2, Reis & Baptista 2016a).

1.1.17 Relação do provérbio com o texto

Tratemos agora das complexas relações que o provérbio pode estabelecer com o contexto discursivo ou comunicativo em que se insere. Segundo Zuluaga (1980) os provérbios são semântica e sintaticamente autónomos em relação ao contexto, funcionando como *um texto dentro de outro texto*. Também Lopes (1992: 37) utiliza igualmente o termo *texto-provérbio* para definir o provérbio como *unidade comunicativa mínima*, isto é:

[...] “um enunciado linguístico completo e coerente, de extensão variável: trata-se, por conseguinte, de uma unidade semântico-pragmática, definida em função do seu potencial comunicativo no quadro de um processo de interacção. Fragmento discursivo globalmente coerente, o texto-provérbio pode ocorrer quer isolado [...] quer integrado numa interacção conversacional quotidiana ou num outro texto mais extenso” [...].

e de modo ainda mais explícito, atribuindo ao provérbio:

[...] “o seu estatuto efectivo de texto linguístico, fragmento discursivo coeso e auto-contido, cuja compreensão se processa segundo mecanismos similares aos que são activados na interpretação dos textos produzidos na interacção coloquial quotidiana”. (*ibidem* 221-222)

Por essa razão, a autora determina dois modos de análise dos provérbios, que podem ser considerados como complementares:

“Tendo em conta este tipo de distribuição, o provérbio pode ser estudado em duas perspectivas complementares: enquanto texto isolado, funcionando como unidade comunicativa mínima, e enquanto texto incorporado num fragmento discursivo mais vasto.” (*ibidem*)

por outras palavras, (i) uma análise interna, que procura identificar as leituras potenciais (lexicais?) que os provérbios autorizam; e (ii) uma análise externa, isto é, na atualização dessas leituras em relação ao texto em que se insere (ou ao contexto comunicativo-pragmático em que são efetivamente usados):

Segundo a autora, o que distingue os enunciados proverbiais das expressões idiomáticas é o facto de os provérbios (1992: 10) “te[re]m *sempre um valor semântico autónomo em termos comunicativos*, ao contrário das expressões idiomáticas que são apenas constituintes de frase e nunca podem ocorrer como enunciados completos” (itálico nosso).

Embora a autora não defina exatamente o que entende por expressões idiomáticas, julgamos que se está a referir àquilo a que chamamos de unidades morfossintáticas ou palavras compostas, de nível intrafrásico, e não de construções do tipo das frases fixas. A própria autora vacila, quando no mesmo fôlego junta sob a designação de locuções idiomáticas expressões fixas “do tipo «bater a bota»” e locuções adverbiais como “de fio a pavio”. Contudo, as frases fixas podem formar enunciados completos. Naturalmente, a referência a constituintes sintáticos vem complicar ainda mais esta terminologia vacilante e ambígua, pois uma frase fixa pode ser argumento de um operador superior, por exemplo numa completiva, e formar, *ipso facto*, um constituinte da frase de ordem superior:

O Pedro pediu à Joana que tirasse nabos da púcara com o João.

Apoiando-se em Taylor, a autora aponta “como traço característico do provérbio o facto de se bastar a si próprio, isto é, de poder funcionar como enunciado completo num acto de comunicação” (*ibidem*). A definição de Taylor, porém, parece ser mais de ordem pragmática: Taylor: “As a guide to life’s problem, the proverb summarizes a situation, passes a judgement, or offers a course of action” (Taylor 1975: 79, *apud* Lopes 1992). A

autora refere ainda a definição de Ageno (1960), que descreve o provérbio como “uma frase autónoma, com significado completo”.

Esta autonomia dos provérbios justifica, então, o facto de serem sentidos pelo locutor como um discurso alheio, como uma citação, frequentemente assinalada quer por mecanismos de introdução no discurso do próprio falante, quer pelo emprego de aspas e outros marcadores gráficos, como os que foram tratados em Reis & Baptista (2016b) e Reis & Baptista (2017b). Estes processos formais não são mutuamente exclusivos e podem combinar-se entre si, veja-se, a esse propósito, o seguinte exemplo:

Como diz o provérbio, ‘não há duas sem três’... e eu sou o terceiro técnico da Roma nesta época¹⁷

Neste caso, o provérbio não só é introduzido pela expressão formulística *como diz o provérbio*, como é ainda assinalado enquanto citação (ou texto alheio) pelo uso de plicas (aqui com o valor de aspas).

Estabelecido o valor do provérbio enquanto *citação* e, nesse sentido, como um *enunciado auto-contido* (Norrick 1985: 32, *apud* Lopes 1992: 10), interessa determinar com algum rigor as relações que esse *texto dentro de outro texto* estabelece com discurso em que está integrado. Foi nesse sentido que procurámos respaldo teórico em diferentes teorias da área por vezes chamada “gramática do texto”.

A Teoria da Estrutura Retórica (Rhetorical Structure Theory – RST), proposta por Mann & Thompson (1988), é provavelmente uma das mais influentes na conceptualização e descrição das relações entre elementos textuais (orações, frases, parágrafos), sendo inclusive a base para diversos trabalhos em processamento computacional do discurso em português (Pardo et al. 2004, Pardo 2005, Maziero 2016), com *corpora* anotados já disponíveis (Cardoso et al. 2011) e servindo para diversas aplicações, nomeadamente a tarefa de sumarização automática (Nóbrega & Pardo 2017).

Esta teoria tem vindo a estudar as relações estabelecidas entre as partes dos textos e na forma como estas estão organizadas para satisfazer o objetivo comunicativo do enunciador. Essas relações podem resultar quer da combinação entre frases ou de porções maiores de texto. Nesta proposta, as frases organizar-se-iam hierarquicamente nos textos de diferentes formas, estabelecendo maioritariamente relações do tipo *núcleo-satélite*, envolvendo uma *oração nuclear* (N) e uma *oração satélite* (S), complementar da primeira, em detrimento das orações multinucleares, que apresentam mais do que um núcleo e nenhum satélite. As relações entre as frases obedecem a quatro tipos de restrições: *restrições do núcleo*; do *satélite*; da *combinação do núcleo e do satélite*; e *efeito* (pretendido pelo escritor e causado no leitor), e podem ser

¹⁷<https://www.record.pt/internacional/detalhe/luigi-del-neri-como-diz-o-proverbio-nao-ha-duas-sem-tres>

de diferentes tipos como, por exemplo, *concessão*, *condição*, *circunstância*, *evidência*, *solução*, *motivação*, *razão*, *sequência*, *elaboração*, *causa* e *contraste*.

Ora, apesar de ser claro o valor do provérbio enquanto “texto alheio citado”, a função que este desempenha em concreto nos textos onde ocorre não é muitas vezes evidente ou é de difícil determinação. Especificamente, este valor de citação que o provérbio implica, parece dificilmente enquadrar-se na proposta teórica da RST, já que (i) a relação núcleo-satélite não parece aplicar-se aqui, ou quanto muito apenas se aplica pelo seu valor de discurso direto reportado, resultante do encaixe da expressão proverbial sob verbos introdutores de discurso (*verba dicendi*); (ii) é difícil fazer corresponder os valores ditos de *efeito* à função que o provérbio efetivamente exerce, quer porque o mesmo provérbio pode, aparentemente, desempenhar várias funções, quer porque há funções que não se aplicam de todo.

1.1.18 Funções discursivas

Um contributo importante, quanto a nós, para esta questão é o trabalho de Lopes (1998). Esta autora analisa um conjunto de provérbios inseridos em textos de imprensa, com o objetivo de determinar as funções discursivas que eles desempenham nesses textos. Verificou que os provérbios que ocorrem no seu *corpus* de trabalho aparecem em diferentes lugares:

- (i) em posição de destaque (por exemplo em títulos);
- (ii) no interior de *leads*;
- (iii) inseridos no corpo de artigos ou crónicas, às vezes funcionando como “remate conclusivo”.

Independentemente da posição em que surgem, eles podem ser citados na íntegra, parcialmente truncados ou mesmo “distorcidos”. Relativamente à distorção dos provérbios, a autora afirma que esta distorção tanto poderá ter uma intenção lúdica como poderá ter uma intenção subversiva, quando parece desconstruir a sabedoria popular.

Do ponto de vista pragmático, a autora identificou as seguintes funções discursivas dos provérbios:

- (a) em títulos e em subtítulos, a função de *resumo catafórico* daquilo que será explicitado depois; nestes casos, ou desempenhando um papel “quase retórico” ou correspondendo a um comentário (de forma indireta) do próprio jornalista;
- (b) no final de artigos ou crónicas, a função de *remate/comentário* ou *sumário anafórico* do texto; e

- (c) no interior de artigos e crónicas, a função de *comentário avaliativo* (do jornalista) ou como peça de um *esquema argumentativo*; neste último caso, o provérbio é geralmente um argumento a favor de uma conclusão ou mesmo justificação de uma afirmação anterior.

Em forma de síntese, a autora agrupa estas funções em duas classes distintas:

- (1) “as que relevam de uma análise das funções ilocutórias dos provérbios inseridos em sequência discursivas”, nomeadamente “as funções de natureza argumentativa e os comentários avaliativos” do autor (*ibidem*: 125);
- (2) “as que relevam de um plano de organização macro-textual”, concretamente, “as funções de sumário anafórico e resumo catafórico”.

Partindo destas observações, procurámos, pela nossa parte, contribuir para a reflexão em torno da *intenção comunicativa* no uso de provérbios no discurso. Esta *intenção comunicativa* liga-se, como é óbvio, à *função* que o provérbio desempenha nos textos em que ocorre, numa determinada situação comunicativa.

Admitimos, com Lopes (1998), que o provérbio possa desempenhar funções pragmáticas diferentes consoante a situação em que é usado. Tal pode fazer variar subtilmente o seu valor, dependendo da intenção comunicativa do locutor. Por outro lado, as funções pragmáticas com que os provérbios são usados podem não se limitar apenas às que, num *corpus* de textos sobretudo de natureza jornalística, esta autora identificou.

Assim, quisemos verificar, sem condicionar à partida a natureza da relação discursiva em jogo, se os falantes conseguiam determinar a intenção de um autor ao usar um provérbio (veja-se questionário 3, *As intenções comunicativas no uso de provérbios* [Anexo 5]), servindo-nos para isso de duas estratégias: por um lado, pedindo aos informantes para identificar os elementos textuais sobre os quais o provérbio é dito; e, por outro lado, pedindo-lhes que avaliassem, numa escala de polaridade (positivo-negativo), qual o valor do provérbio relativamente aos elementos do texto sobre os quais ele é dito. Desta investigação e dos seus resultados, damos conta em Reis & Baptista (2019) [corresponde ao Artigo 10 desta tese].

Nesta primeira parte da introdução, procurámos identificar algumas das temáticas pertinentes, bem como delimitar os conceitos-chave e diferentes abordagens aos fenómenos complexos e multímodos dos provérbios. Na secção seguinte, iremos apresentar de forma mais sistemática os objetivos deste estudo, as questões de investigação que o justificam, uma visão panorâmica dos métodos utilizados. Ao longo desta secção indicaremos os trabalhos que desenvolvemos ao abordar cada um dos diferentes tópicos e que se encontram coligidos na segunda parte da tese.

1.1.19 Objetivos e métodos

Neste trabalho, temos como primeiro objetivo contribuir para um melhor conhecimento dos provérbios do português europeu, dos seus usos e da variação formal que apresentam, melhorando a identificação automática deste tipo de expressões em textos. Os resultados deste estudo poderão, num segundo momento, servir para diferentes aplicações, para as quais demos igualmente os primeiros passos.

Na secção anterior desta introdução, apresentámos alguns tópicos relevantes para a descrição linguística dos provérbios, identificando aspetos em que parece haver necessidade de aprofundamento dos conhecimentos atuais sobre este tipo de expressões.

Assim, este estudo é norteado pelas seguintes questões de investigação, a que pretendemos dar resposta ao longo da pesquisa:

- (1) Que metodologias serão mais eficientes para detetar os provérbios automaticamente em textos, sobretudo quando estes não estão associados a mecanismos formais que os introduzem no discurso?
- (2) Que correlação é possível estabelecer entre a frequência de uso de provérbios em textos e a disponibilidade lexical de provérbios e variantes?
- (3) Que relações estabelece o provérbio com o discurso em que se insere? Como é possível determinar essas relações? Por outras palavras, como explicar o uso e o significado dos provérbios nos contextos comunicativos em que ocorrem?

Relativamente à primeira questão, verifica-se que a maior parte dos sistemas de processamento computacional de textos desenvolvidos para o português (e.g., PALAVRAS¹⁸: Bick 2000; LX-Parser¹⁹: Silva et al. 2010) ignoram, não identificam ou não dão adequadamente conta das expressões proverbiais, mesmo as mais usuais. Por outro lado, a variação formal interna dos provérbios torna não-trivial a tarefa de os reconhecer nos textos em que ocorrem. Será necessário determinar métodos eficientes de os representar e à sua variação, reconhecendo a unicidade subjacente a essa variação (as unidades paremiológicas), por forma a poderem ser implementados nesse tipo de sistemas de processamento de língua natural. Tais métodos deverão, além disso, garantir a reprodutibilidade das observações realizadas, bem como a cumulatividade dos dados coligidos.

Quanto à segunda questão de investigação, a frequência é, sem dúvida, um parâmetro relevante mas não o único na determinação de um “núcleo duro” de provérbios usuais da língua. Este conjunto de expressões virá a constituir um *mínimo paremiológico* do português, reunindo as unidades paremiológicas (provérbios

¹⁸https://visl.sdu.dk/constraint_grammar.html

¹⁹<http://lxcenter.di.fc.ul.pt/tools/pt/conteudo/LXParser.html>

e respetivas variantes), sobre as quais se poderão, então, concentrar os esforços de descrição lexicográfica (para fins de processamento computacional da língua, por exemplo) e que poderão nortear a seleção de materiais para diferentes aplicações (didáticas, terapêuticas, entre outras). Tanto quanto sabemos, este trabalho de constituição do mínimo paremiológico ainda não foi feito para o português, pelo menos de uma forma que corresponda ao dos *standards* internacionais.

Finalmente, a questão de investigação das relações retóricas que o provérbio estabelece com o texto em que ocorre e a sua função discursiva é um tema complexo, já identificado na literatura, mas que merece um estudo mais aprofundado. De facto, o estabelecimento destas relações invoca categorias que, provavelmente, só são ainda parcialmente conhecidas ou foram apenas brevemente descritas (Lopes 1992, 1998), parecendo resistir ainda a um enquadramento adequado em abordagens teóricas atuais e populares da tarefa de análise (automática) de discurso, como a RST (Mann & Thompson 1988).

São, portanto, objetivos específicos desta tese:

- (i) contribuir para a identificação automática de provérbios em textos, com especial referência aos provérbios da variante europeia do português;
- (ii) estabelecer uma tipologia formal (sintática) para os provérbios do português;
- (iii) estudar a relação complexa, a nível discursivo, entre os provérbios e o texto em que se encontram;
- (iv) determinar um conjunto de provérbios muito usuais e conhecidos da maioria da comunidade linguística de falantes do português europeu;
- (v) contribuir para o desenvolvimento de jogos didáticos para a aprendizagem de português, construídos a partir de provérbios usuais, tendo como horizonte métodos de construção automática desses jogos;
- (vi) contribuir para o desenvolvimento de ferramentas de apoio ao diagnóstico/terapia para diversas afasias que recorram aos provérbios como objeto linguístico de referência (estímulo/alvo).

Para atingir vários destes objetivos, será necessário, num primeiro momento, desenvolver métodos para identificar automaticamente as expressões proverbiais nos textos onde ocorrem, com recurso a ferramentas do processamento da linguagem natural (PLN). Tal não é certamente uma tarefa fácil se considerarmos a variação formal que estas estruturas apresentam e a forma como se integram no discurso. Um outro fator a considerar é o facto de não ser inteiramente consensual o que se considera por um provérbio, sendo assim imperativo começar por estabelecer fronteiras formais que permitam delimitar esse objeto de estudo (v. Artigo 7).

Para identificarmos os provérbios nos textos, partimos da tipologia formal (sintático-semântica) proposta por Rassi, Baptista, & Vale (2014) para os provérbios do português do Brasil. Desenvolvemos a nossa própria proposta de tipologia (v. Artigo 1 e Artigo 8), tendo como propósito permitir que esta oriente o desenvolvimento de processos de identificação automática deste tipo de estruturas. Em perspectiva está, para já, o desenvolvimento de um procedimento formal automático, com recurso a transdutores de estados finitos (ou *grafos*) (Paumier 2014), que permita a classificação da extensa base de dados de provérbios por nós coligida (Reis et al. 2019)²⁰.

De facto, considerámos necessário desenvolver primeiramente uma base de dados de provérbios para o desenvolvimento do estudo. A base de dados, nesta sua primeira forma, não só nos serviu como fonte para podermos determinar as diferentes estruturas sintáticas dos provérbios, como também nos serviu como *corpus* de teste para avaliar as ferramentas de identificação que foram sendo desenvolvidas ao longo do estudo. Num momento inicial, a base de dados era constituída por cerca de 56.000 provérbios (e variantes). Foi compilada a partir da digitalização de duas coletâneas de provérbios (Costa 1999, Machado 2011). Mais tarde, foram digitalizadas mais duas coletâneas de provérbios (Moreira 1996, Parente 2005), o que permitiu reunir um total de 114.413 entradas (v. Artigos 2 e 3). Usou-se, então, um OCR (*Optical Character Recognition*). Considerámos essencial procedermos à avaliação deste OCR, de modo a podermos determinar a qualidade do texto digitalizado, e assim descartar a hipótese de os provérbios não serem capturados pelas ferramentas desenvolvidas devido a erros de digitalização do texto. Os resultados obtidos nesta avaliação, feita por amostragem, mostraram uma taxa de erro de 0,0047, sobre o total de 54.747 palavras. Os dados dessa avaliação estão disponíveis no Anexo 3 desta tese.

As ferramentas produzidas para a identificação automática de provérbios em textos foram sendo adaptadas de acordo com os resultados obtidos nas diferentes fases deste trabalho. Em cada artigo será minuciosamente apresentada a metodologia utilizada. De uma forma muito geral, podemos antecipar que foram utilizadas (pelo menos) duas abordagens diferentes para esse fim: uma com recurso a transdutores de estados finitos (ou grafos; v. Artigo 4; ver Anexo 4) e uma outra com recurso a expressões regulares (v. Artigo 3).

São parte desta primeira abordagem:

- (a) os transdutores construídos semiautomaticamente com base nas palavra-chave dos provérbios e que aceitam algumas inserções (sinais de pontuação ou palavras) entre estas;

²⁰Comunicação em conferência.

Disponível em: https://www.researchgate.net/publication/336085113_Automatic_Syntactic_Classification_of_European_Portuguese_Proverbs

- (b) os transdutores construídos manualmente, que representam unidades paremiológicas e que aceitam variação morfossintática dos constituintes e inserções entre palavras;
- (c) os transdutores que representam:
 - (i) as expressões introdutórias muitas vezes associadas ao uso dos provérbios (e.g. *como diz o provérbio*),
 - (ii) os termos relacionados com o conceito de *provérbio* (e.g. *rifão*, *adágio*).

Realçamos que as palavras-chave de cada provérbio foram determinadas ao ser aplicado à base de dados de provérbios um filtro com uma lista de *stopwords* (ver Anexo 1), por nós estabelecida. As *stopwords* correspondem essencialmente a palavras gramaticais (e.g. determinantes, pronomes, preposições) que julgamos não contribuírem de forma essencial para a tarefa de identificação dos provérbios. Este procedimento permitiu simplificar as listagens iniciais e agrupar (semiautomaticamente) os provérbios compilados de acordo com as palavras-chave que os constituem. Por exemplo, o provérbio *À sombra da noqueira não te deites a dormir* tem como palavras-chave [sombra noqueira deites dormir]. Uma vez que há vários provérbios que apresentam as mesmas palavras-chave, este procedimento permitiu-nos reduzir a lista dos mais de 114.000 provérbios para cerca de 52.000 sequências de palavras-chave diferentes.

Uma outra abordagem consistiu na construção de comandos **grep** com expressões regulares para a pesquisa de sequências de duas a cinco palavras-chave, aceitando algumas inserções entre elas.

Numa última fase deste trabalho, foram desenvolvidas expressões regulares para representar as 318 UP do mínimo paremiológico, que foram em seguida integradas como gramáticas locais na cadeia de processamento de língua natural STRING (Mamede et al. 2012)²¹. Estas gramáticas locais permitem à STRING processar textos, identificando e delimitando os provérbios usuais que neles ocorrem.

As ferramentas de identificação automática de provérbios em textos serão, pois, utilizadas para a determinação dos provérbios mais conhecidos e mais frequentemente usados no português (cf. Artigos 5 e 9), outro objetivo do nosso trabalho. Dito de outra forma, usamos essas ferramentas para o estabelecimento de índices de frequência e distribuição dos provérbios numa variedade de *corpora* de diferentes tipos textuais (cf. Artigos 3, 4 e 5), a fim de permitir determinar quais são, efetivamente, os provérbios que os falantes conhecem e usam habitualmente. Assim, por exemplo, comparou-se a distribuição dos provérbios nas coletâneas por nós compiladas, bem como a sua frequência na *web*.

²¹<https://string.hlt.inesc-id.pt/demo>

Paralelamente a essas estratégias, recorrer-se-á a questionários (Anexo 5), que não só visam aferir se os falantes conhecem e usam um conjunto de provérbios selecionados aleatoriamente (ver o questionário *Provérbios Usuais*); ou se conseguem completar corretamente um conjunto de jogos com provérbios (*quizzes*), construídos de acordo com critérios linguísticos bem motivados (ver o *Questionário de avaliação da qualidade de jogos proverbiais*).

Ao determinarmos a disponibilidade lexical dos provérbios, estabelecemos o conjunto dos provérbios mais frequentemente usados na língua e melhor conhecidos pela maioria dos falantes: é a este conjunto que chamamos o mínimo paremiológico (ver Anexo 6). Com este instrumento, esperamos contribuir para a elaboração de instrumentos de diagnóstico e terapêutica fiáveis (ver os jogos desenvolvidos nos Anexos 7 e 8), por permitirem uma seleção mais rigorosa e informada dos exemplos que funcionam como estímulo e, até, ajudando a melhor adequar esses testes ao perfil sociológico/clínico dos sujeitos (v. Artigo 6).

Da mesma forma, esse mesmo conjunto de provérbios usuais da língua pode contribuir para que os materiais a serem integrados nos manuais de ensino de português – quer enquanto língua materna, quer enquanto língua estrangeira – (ou mesmo nas próprias recolhas de provérbios, dependendo do público-alvo), sejam mais representativos da dimensão cultural da língua. Efetivamente, verificámos que a presença de expressões proverbiais em manuais de português língua materna, é, por vezes, pouco motivada (ver Artigo 3). Já em relação aos manuais de português língua não materna (Artigo 4), verificámos que já apresentam, na sua maioria, os provérbios mais usuais. Esses provérbios poderão ainda ser utilizados para a construção de jogos didáticos, que sirvam de treino de diferentes fenómenos da língua, por exemplo, jogos que visem o estudo de fenómenos morfossintáticos, semânticos ou outros (e.g. rimas, aliterações) (veja-se Anexo 7).

1.1.20 Estrutura da tese

Na Parte I desta tese foi apresentada uma introdução que procurou abordar, de forma muito geral, as principais temáticas desenvolvidas ao longo desta investigação.

Na Parte II apresenta-se o conjunto de artigos já publicados (7 artigos) ou aceites para publicação (3 artigos).

A ordem pela qual os artigos são apresentados parece-nos ser a mais coerente para uma melhor compreensão do trabalho desenvolvido, não correspondendo à ordem da sua publicação. Assim, estes artigos apresentam-se pela seguinte ordem:

Artigo 1 – Portuguese Proverbs: Types and variants;

Artigo 2 – Let's play with proverbs? – NLP tools and resources for iCALL

applications around proverbs for PFL;

Artigo 3 – O uso de provérbios no ensino de português;

Artigo 4 – Os provérbios em manuais de português língua não materna;

Artigo 5 – Estimating Lexical Availability of European Portuguese Proverbs;

Artigo 6 – O provérbio como estímulo num terapeuta virtual;

Artigo 7 – Estudo das estruturas proverbiais com sujeito indefinido – Provérbio ou expressão fixa?;

Artigo 8 – Descrição e classificação sintática das expressões proverbiais do português europeu;

Artigo 9 – Determinação de um mínimo paremiológico do português europeu; e

Artigo 10 – As intenções comunicativas no uso de provérbios.

A tese termina com uma conclusão, onde são realçados os aspetos mais importantes desenvolvidos, as principais constatações a que chegamos, bem como os resultados obtidos nesta investigação.

Finalmente, apresentam-se as referências bibliográficas e os anexos.

O Anexo 1 apresenta a lista das *stopwords* aplicada à base de dados; no Anexo 2, são apresentados os critérios para agrupar variantes de provérbios sob uma mesma unidade paremiológica; no Anexo 3, são apresentados os resultados da avaliação do OCR (Optical Character Recognition) que nos serviu para digitalizar as coletâneas de provérbios da nossa base de dados; no Anexo 4, apresenta-se os grafos desenvolvidos para cada Unidade Paremiológica; no Anexo 5, podem ver-se os três questionários desenvolvidos: (1) *Provérbios Usuais*, (2) *Questionário de avaliação da qualidade de jogos proverbiais* e (3) *As intenções comunicativas no uso de provérbios*; no Anexo 6, está representado o *Mínimo paremiológico do português europeu*; no Anexo 7 apresenta-se uma primeira lista de jogos com provérbios que poderá ser utilizada quer para o desenvolvimento de instrumentos de diagnóstico ou terapia de certas patologias da linguagem, quer para a aprendizagem de português; o Anexo 8, contem os jogos desenvolvidos com o objetivo de virem a ser integrados na plataforma VITHEA-Kids (Mendonça et al. 2015); e finalmente os Anexos 9 e 10 apresentam o conjunto de provérbios selecionados pelos anotadores como ‘usuais’, retirados da lista dos 114.413 provérbios da base de dados, e que serviram de base, juntamente com outros critérios, para a determinação do mínimo paremiológico do português europeu.

Parte II

Publicações

Nesta parte da tese, apresentamos o conjunto de trabalhos desenvolvidos durante o período de investigação no quadro do doutoramento em Ciências da Linguagem.

A ordem pela qual os trabalhos são apresentados não corresponde à ordem pela qual foram publicados. Considerou-se que esta organização era a mais coerente e a que mais contribuía para uma melhor compreensão da investigação desenvolvida.

Os trabalhos apresentados nesta parte reproduzem exatamente as publicações e seguem a formatação exigida pelos respetivos editores. Por isso, esta parte apresenta dupla paginação, uma vez que se preservou a paginação original onde os trabalhos foram publicados, tendo-se colocado também, no canto superior direito, o número de página desta tese.

O conjunto de textos apresentado é da autoria da candidata. O orientador da tese figura, naturalmente, como co-autor dos trabalhos publicados, mas a responsabilidade e eventual mérito dos trabalhos é apenas da autora destas linhas. O orientador participou ainda no processo de anotação conjunta do *corpus*.

No caso do artigo 6 – *O provérbio como estímulo num terapeuta virtual* – apresenta-se um conjunto de jogos desenvolvidos com base em provérbios para a sua inclusão da plataforma VITHEA, um terapeuta virtual para o tratamento de afasias. Neste estudo, a descrição da plataforma foi feita pelos respetivos autores, Anna Pompili e Alberto Abad.

Capítulo 1

Portuguese Proverbs: Types and Variants

Artigo 1

Publicado em: Gloria Corpas Pastor (Ed). *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives*. Geneva: Editions Tradulex, pp. 208-217

PORTUGUESE PROVERBS: TYPES AND VARIANTS

Sónia Reis

University of Algarve
reis.soniamm@gmail.com

Jorge Baptista

University of Algarve
L2F/INESC-ID Lisboa
jbaptis@ualg.pt

Keywords: Proverbs, European Portuguese, Variation, Automatic Identification, Corpus Linguistics

Abstract

Drawing on the methodology and previous results of Rassi *et al.* (2014) on the automatic identification of Brazilian Portuguese proverbs, this paper reports on an extension of that experiment, but now focused on the identification of the European Portuguese proverbs and their variants. Based on a large collection of over 56 thousand Portuguese proverbs and their variants, a database of proverb types was specifically built for natural language processing, along with the finite-state tools that allow for the identification of these strings in texts. Our aim is to make these linguistic resources and language processing tools publicly available, which will undoubtedly be deemed useful assets to other paremiologic studies.

1. INTRODUCTION

Proverbs are an important part of most societies' culture and language. As micro-texts, brought into discourse from the common cultural repository, they are subject to many creative types of variation. On the other hand, functioning as in quotation mode, they integrate discourse in an almost disruptive way, challenging natural language processing (NLP) systems, and requiring their accurate identification and delimitation.

Concerning Portuguese proverbs, and though several, extensive collections of proverbs are available in printed form (Machado, 2011), to the best of our knowledge, no resources have been specifically produced for NLP purposes, even if some digitally available dictionaries (Almeida, 2014) include a few examples, interspersed between other type expressions, like different types of idioms and many forms of slang.

Recently, Rassi *et al.* (2014) have proposed a formal (syntactic) classification of proverbs, based on a collection of over 3,500 proverb variants, organized in 594 proverb types (Rassi, 2014), and taken from several dictionaries from the Brazilian variety of the Portuguese language. The authors presented a finite-state based method for the automatic identification of proverbs in large-sized corpora, and experimented on a 29M tokens corpus of journalistic text (Bruckschein *et al.* 2008), taken from the daily online edition of the Brazilian newspaper *Folha de São Paulo*. The authors report a 60 to 73% precision, depending on the proverb class and the width of the insertion window between the proverbs' keywords. In spite of the corpus size, but not surprisingly, only 137 types and

788 instances were matched, most likely because of the journalistic nature of the texts in this corpus. However, seen from this side of the Atlantic, results from Rassi and colleagues are surprising mostly for the fact that, in spite of some American idiosyncrasies, most proverbs seem to exist also in the European variety, quite unlike the mismatch that has been found for verbal idioms (undisclosed reference).

Drawing on this methodology and these previous results, this paper reports on an extension of that experiment, but now focused on the identification of the European Portuguese proverbs and their variants. This intends to set up the basis for a large collection of Portuguese proverbs and their variants, specifically built for natural language processing, and to make it publicly available, along with the finite-state tools built for retrieving them from texts. These tools and resources will undoubtedly be deemed useful assets to other paremiology studies.

The remainder of this paper is structured as follows: Section 2 presents the methods, starting with the formal classification criteria, and the current state of the collection of Portuguese proverbs' types (§2.1); followed by the criteria to select the proverbs' keywords in order to define those proverbs' types (§2.2); next, the finite-state tools to match proverbs in text are presented (§2.3), followed by the proverbs' collections used to produce a digitized list of Portuguese proverbs and variants (§2.4). Section 3 present some preliminary evaluation of the finite-state tools when applied to this list and discusses some of the issues pertaining to the resulting matches. Finally, Section 4 concludes the paper and hints at future work.

2. METHODS

2.1 Adapting the proverbs classification of Rassi *et al.* (2014) to European Portuguese

Based on the syntactic classification of Brazilian Portuguese proverbs proposed by Rassi *et al.* (2014) we produced a database, in tabular format, with the key elements of each proverb. This formal, taxonomic, approach allows us to determine accurately the concept of variant and base form of a proverb.

In order to better frame this paper, we first present the formal/syntactic classification proposed by Rassi *et al.* (2014) of Brazilian Portuguese proverbs, which may also be adopted by and large to European Portuguese proverbs.

This classification is based on the number of propositions forming the proverb and the part-of-speech (PoS) of their main elements. Other, secondary features are also used. The number of propositions (or clauses) organizes the data in 3 main classes (**P1x**, **P2x** and **P3**). The specific sentence type of **P1x**, or the transformations it may (or may not) undergo are used to further split this type into several classes. Thus, we find the following classes:

P1F1: impersonal constructions; while in European Portuguese this are mostly sentences with the verb *haver* (there be), in Brazilian Portuguese one also finds an impersonal use of *ter* (have); the head noun is often modified by a prepositional phrase (PP) or even a relative subclause; impersonal constructions with indefinite clitic pronoun *se* were also included in this class:

Não há rosa sem espinhos ‘There is no rose without thorns’

Não há mal que sempre dure ‘There is no evil that lasts forever’;

Tem muita estrela pra pouco céu ‘There is ~~to~~^{too} many star[s] for little/small sky’

Devagar se vai ao longe ‘Slowly one can go far’

P1F2: attributive sentences with copula verb *ser*; the subject is usually a noun (sometimes a verb in the impersonal infinitive); the predicative element can be a noun, an adjective, an verb in the impersonal infinitive, or even a prepositional phrase:

A fome é o melhor tempero ‘Hunger is the best seasoning’

O amor é cego ‘Love is blind’

Partir é morrer um pouco ‘To leave is to die a little’

O silêncio é de ouro ‘Silence is of gold’

P1F3: direct-transitive verb constructions:

Muitos cozinheiros estragam a sopa ‘[Too] many cooks ruin the soup’

Os fins justificam os meios ‘The ends justify the means’

Uma mão lava a outra ‘One hand washes the other’

P1F4: obligatory negation:

Burro velho não aprende línguas ‘Old donkey does not learn languages’

Uma andorinha não faz a primavera ‘A single swallow does not make spring’

Gostos não se discutem ‘Tastes are not discussed’

P1F5: obligatory fronting of the prepositional complement:

De boas intenções está o inferno cheio ‘Hell is full of good intentions’

Em terra de cegos quem tem um olho é rei ‘In a land of blind [people], he who has an eye/the one-eyed is king’

Em boca fechada não entra mosca ‘In [a] closed mouth no fly enters’

Next, two-clause proverbs are considered (**P2x** classes). These include:

P2F1: proverbs with a main clause and a comparative subordinate clause; as comparison can be expressed in many different ways, different types of comparative structures are considered:

Antes tarde (do) que nunca ‘Better late than never’

Mais vale um pássaro na mão do que dois a voar ‘Better a bird in the hand than two fling’

Não há pior cego do que aquele que não quer ver ‘There is no one more blind than the one who does not want to see’

P2F2: proverbs with two coordinate clauses; in some cases, the coordinative conjunction is not expressed but only implied (asyndeton):

As moscas mudam mas a merda é sempre a mesma ‘The flies change but the sheet is always the same’

Vão-se os anéis [mas, e] ficam os dedos ‘The rings go away (but/and) the fingers stay’

Deus não fecha uma porta que não abra logo duas ‘God does not close a door without opening two right away’

P2F3: verb-less proverbs with two phrases; in some cases, a coordination or subordination nexus can be inferred but the coordinate/subordinate conjunction is not expressed,

only implied (asyndeton); in other cases, the (implied) main verb of the sentence can be inferred:

Cada roca com seu fuso, cada terra com seu uso ‘Each spinning distaff with its spindle, each land with its ways

Muito riso, pouco siso ‘[Too] much laughter too little judgement’

Tal pai, tal filho ‘Like father, like son’

Olho por olho, dente por dente ‘Eye for eye, tooth for a tooth’

Cada cabeça [?dá] sua sentença ‘Each head [gives] its sentence (opinion)’

P2F4: pseudo-interrogative sub-clauses, introduced by interrogative *Qu-* (*Wh-*) pro-forms:

O que não mata engorda ‘What doesn’t kill [one] makes [one] fat’

Quem avisa amigo é ‘He who warns [one] is a friend’

P2F5: proverbs with a main clause and a subordinate clause:

Fazer o bem sem olhar a quem ‘To do good without looking to whom’

Devagar que tenho pressa ‘Slowly for I am in a hurry’

P2F6: proverbs with a main clause and an obligatorily fronted subordinate clause:

Para morrer basta estar vivo ‘In order to die, is enough to be alive’

Enquanto houver vida há esperança ‘While there is life, there is hope’

Quando a esmola é muita, o pobre desconfia ‘When the charity is too much, the poor gets suspicious’

Finally, class **P3** includes the long proverbs with more than 2 clauses/propositions. Unlike the previous classes, this one has not been subdivided yet, pending on a accumulation of data that would render such sub-classification necessary. Thus, in **P3** we find:

P2F6: proverbs with more than 2 clauses/propositions; the coordinative conjunction can be omitted; often, instead of a clause one finds verb-less phrases:

Mãos frias, coração quente, amor ardente ‘Cold hands, hot heart, burning love’

Laranja, de manhã é ouro, à tarde é prata e à noite mata ‘Orange, in the morning is gold, in the afternoon is silver and at night it kills’

Um é pouco, dois é bom, três é demais ‘One is [too] little, two is good, three is too much’

Following their publication, A. Rassi and her co-authors published the list of 594 proverb types and their classification⁴⁴, which we used as the basis for our own classification of European Portuguese proverbs. While, in its general features, their taxonomy seems a useful tool to organize the complex and abundant data already available, in some cases, we disagreed with the authors’ classification and decided to assign some of their proverbs to a different class. We also some added Portuguese-specific proverbs and variants, absent from Rassi et al. (2014) list. This careful and close review of the proverbs list, along with the growing number of classified proverbs, will eventually lead to a more granular classification, especially for the larger classes, that is, those with the larger number of types.

⁴⁴ https://www.researchgate.net/publication/266852580_Rassi2014?ev=prf_pub.

For instance, the proverbs with the copula verb *ser* (be) constitute such a large set that it could be advantageous to create sub-classes according to the morphological class (PoS) of the element that are associated (adjective, noun, etc.). Table 1 shows the classification in its current state:

Class	Structure	definition	Example/gloss (or wbw translation)	types	%
P1F1	0 V w	impersonal constructions	<i>Não há rosa sem espinhos.</i> 'There is no rose without thorns'	20	0.03
P1F2	N ₀ Vcop (N+Adj)	predicative constr. (copula verb)	<i>O amor é cego.</i> 'Love is blind'	53	0.09
P1F3	N ₀ V N ₁	direct transitive (no PP)	<i>Uma mão lava a outra.</i> 'One hand washes the other'	80	0.13
P1F4	N ₀ Neg V w	obligatory negation	<i>Uma andorinha não faz a primavera.</i> 'a single sparrow does not makes spring'	53	0.09
P1F5	Prep N ₁ , N ₀ V	obligatory fronting of PP	<i>Pela boca morre o peixe.</i> 'By the mouth dies the fish'	45	0.08
P2F1	F ₁ Cs-comp F ₂	comparative	<i>Mais vale um pássaro na mão do que dois a voar.</i> 'Better a bird in the hand than two flying'	39	0.07
P2F2	F ₁ (Cc) F ₂	coordinate (asyndeton)	<i>Vão-se os anéis, ficam-se os dedos.</i> 'The rings go away, the fingers stay'	71	0.12
P2F3	N ₁ (Cc) N ₂	coordinate (w/o verb)	<i>Tal pai, tal filho.</i> 'Like father, like son'	48	0.08
P2F4	Quem V V w	interrogative subject Qu- 'Wh-'	<i>Quem vê caras não vê corações.</i> 'Who sees faces does not see hearts'	90	0.15
P2F5	F ₁ Cs F ₂	subordinate	<i>Fazer o bem sem olhar a quem.</i> 'To do good without looking to whom'	20	0.03
P2F6	Cs F ₂ , F ₁	obligatory fronting of subordinate	<i>Quando um burro fala, os outros abaixam as orelhas.</i> 'When a dunkey speaks, the others lower their ears'	28	0.05
P3	F ₁ C F ₂ C F ₃	3-clause	<i>Mãos frias, coração quente, amor ardente.</i> 'Cold hands, warm heart, burning love'	47	0,08
				594	

Table 1. Classification of Portuguese proverbs (adapted from Rassi *et al.* 2014).

2.2. Compiling the proverbs' keywords

For each class, we defined the key elements that were to be matched in order to unambiguously identify the proverb. As Rassi and co-authors had not published this data (only an example per type and its class are provided), we have re-done most of their work, carefully reviewing the selection criteria that help define the proverb's core elements. These keywords vary depending on the class and, in some cases, even on the proverb itself and its variants. Therefore, only a glimpse of the complex process of selecting the keywords can be provided here.

For keywords, the main content words (nouns, verbs and adjectives) are usually selected, and represented by their lemma (represented inside chevrons, '<' and '>'), rather than the surface (inflected) form, in order to allow for the capture of creative reuse of the proverb. Hence in, for the entry

[P1F3] *Os fins (não) justificam os meios* 'The ends (do not) justify the means'

the lemmas of the two nouns, *fim* and *meio*, and the verb are considered keywords:

<fim> <justificar> <meio>

Usually, copula verbs and auxiliaries are dropped, as in:

[P3] *Um é pouco, dois é bom, três é demais*

'One is too few, two is good, three is too much'

where only the subjects and the adjectives are kept; notice that in this case, the numerals are not determining an noun, as determinants are usually discarded; the specific word order is characteristic of the proverbial nature of the sentence:

<um> <pouco> <dois> <bom> <três> <demais>

In some cases, it is the structure, rather than the specific words, that is key to the proverb. For example, in the proverb:

[P2F1] ***Duas cabeças pensam melhor do que uma*** ‘two heads think better than one’

we can group the many variants using the following string⁴⁵:

<dois> <N;p> <V;p> (mais+melhor) (do+<E>) que <um:s>

where <N;p> and <V;p> stand for any noun and verb (in the plural), the comparative adverbs *mais* ‘more’ and *melhor* ‘better’ introduce the subordinate comparative conjunction (*do*) ‘than’, which allows for the zeroing of *do*, and the numeral *um* ‘one’.

The number of variants of a proverb can yield quite complex expressions, as in the next case (all variants taken from proverbs’ collections; see §2.3, below):

[P1F4] ***Não fales de corda em casa de enforcado***

Não se deve falar em corda em casa de enforcado

Não se fala em corda em casa de enforcado.

Não se fale em corda em casa de enforcado

É falar de corda em casa de enforcado

Em casa de enforcado não nomeies o baraço

Em casa de enforcado não se fala de corda

Em casa de ladrão não fales em baraço

Em casa de ladrão não lembrar baraço

all of them meaning approximately the same: ‘Do not speak of rope in a hanged man’s house’. Besides the fronting of the prepositional phrase (a locative), in the last for examples, notice the alternation between imperative and the use of the modal auxiliary *dever* ‘should’ (second sentence), the lexical variation of *corda* ‘rope’ and *baraço* ‘string’, the surprising alternation between *enforcado* ‘hanged man’ and *ladrão* ‘thief’, and the alternative use of *lembrar* ‘remember’ instead of *falar* ‘speak’. Notice also the (truncated?) form, in the fifth line, introduced by *ser* ‘to be’, which can be used as a comparison and adapted for commenting on an given event/situation: *Fazer isso é (como) falar de corda ...* ‘to do this is like...’. In this more complex case, and because of the changes in word order, the key elements are represented by two strings:

(não+<E>) (<falar>+<lembrar>) (<corda>+<baraço>) <casa>(<enforcado>+<ladrão>)
 <casa> (<enforcado>+<ladrão>) (não+<E>) (<falar>+<lembrar>) (<corda>+<baraço>)

Notice that in the case of the negation adverb *não* ‘not’, this should be treated as a facultative element, since the zeroing of the negation is often one of the strategies to creatively adapt the proverb to new uses. This very phenomenon is illustrated in one of the proverbs’ variants.

Thus, a database in tabular format was produced, where the lines contain the proverbs keywords and the number of columns varies according to the class. This matrix, that we call a *lexicon-grammar*, can also be used to include other relevant information. For the moment, this is limited to the proverbs’ conventional ID (the code of the class and the proverb number).

⁴⁵ We use the notation of the UNITEX system (Paumier 2003, 2014) in representing these regular expressions: elements inside brackets and linked by ‘+’ are interchangeable in that given position; <E> stands for the null string.

proverb

2.3. Building finite-state tools for ~~adverb~~ identification

Since the lexicon-grammar of proverbs is in constant update, and can not be directly used to match strings in texts, we used UNITE⁴⁶ linguistic development platform (Paumier 2003, 2014), which is based on finite-state technology, we built the tools that would allow to find, delimit and tag as proverbs candidate strings in running text. In this way, the linguistic information is represented independently of the tools used to apply it to texts. Next we describe this process.

A reference graph was created for each class of proverb, according to its syntactic structure and the number of keywords. Each graph describes the sequences of those key elements of the proverbs. In the graph, variables represent the corresponding cells in the matrix. Then, the reference graph is intersected with the matrix: the system reads each line of the database at a time, replacing the variables for the content of the corresponding cells in the matrix and generates a sub-graph for each proverb. Each sub-graph is univocally numbered and entire set of is grouped together in a result graph that can be used to match proverbs in texts. To simplify the procedure, for this paper, a single reference graph was produced and a simplified extract is shown in Fig. 1:

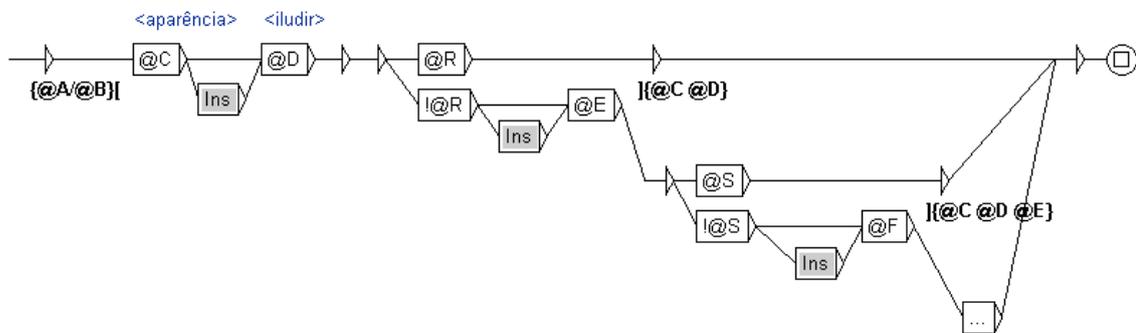


Fig. 1. Reference graph for the proverbs' database.

In this graph, variables @R and @S stand for supplementary or auxiliary columns in the matrix. These columns indicate whether the $n+1^{th}$ column has content (and then this is marked with a plus '-') or not ('+'), the n^{th} column being the last content cell that has been processed; up to 11 columns were used to represent the keywords of the proverbs; in this case, @R stands for the 3rd column, @S for the 4th, and so on; there are always at least two keywords/elements, as illustrated by the proverbial expression *As aparências iludem* 'Appearances deceive'. The system replaces the content variables in the graph (in this case, @C, @D, @E and @F) by their corresponding elements in the matrix. When reaching a variable for the auxiliary columns, the system either continues the path, if this corresponds to a plus sign '+', ending the representation of the proverb; or it moves to the next content variable. The process is repeated until all columns have been explored. Thus, a single reference graph can be used for tables with any given number of columns. The sub-graphs are transducers that can be applied to texts. When matching a given sequence, they insert a right delimiter ']' preceded by the proverb ID and the number of keywords (variables @A and @B); and a left delimiter '[' followed by the proverb's keywords (variables @C, @D, etc.). An insertion window from 0 up to 3 words is represented by an auxiliary sub-graph **Ins** (shown in a grey box). Fig. 2 shows the sub-graph automatically generated for proverb *Roma e Pavia não se fizeram num dia* 'Rome and Pavia were not made in a single day' [P1F4]:

⁴⁶ <http://www-igm.univ-mlv.fr/~unitex/>

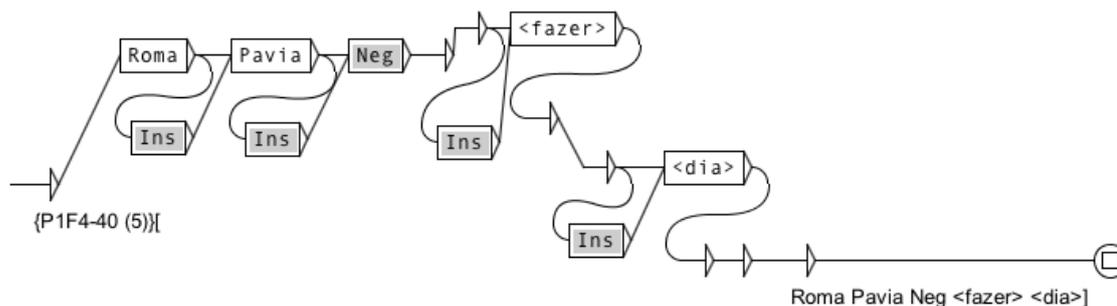


Fig. 2. Sub-graph describing an individual proverb.

2.4. Corpus building

For the automatic identification of proverbs in texts, we digitized a list of approximately 56,150 proverbs (and variants), taken from two large proverb collections:

- Costa (1999): 32,709 entries and 278,217 words;
- Machado (2011): 23,437 entries and 211,257 words.

This list will function not only as a source for the completion of the lexicon-grammar, but also as a testing ground for the finite-state tools built to identify proverbs in texts.

3. PRELIMINARY EVALUATION

The 594 sub-graphs (one for each proverb type), resulting from the process described in the previous section, were applied to the list of 56,150 proverbs (and variants) taken from the two large proverb collections mentioned above. Table 2 shows some preliminary results.

Number of proverbs	LPP (Costa, 1999)	GLP (Machado, 2011)
Proverbs in each <i>corpus</i> (matches)	# 446	# 391
Different proverbs in each <i>corpus</i> (types)	# 199	# 194
Proverbs in both <i>corpora</i> (matches)	# 347	# 310
Different proverbs in both <i>corpora</i> (types)	142	

Table 2. Number of proverbs in corpora

For the 594 proverb types represented in the graphs, only 446 and 391 were captured in each sub-corpus. Notice that this difference may be related to the fact that the list from Machado (2011) is substantially shorter than the one from Costa (1999). Notice also that in spite of the number of matches, the number of different types is much smaller, corresponding to a ratio variants/type of 2.24 in Costa (1999) and 2.02 in Machado (2011). On the other hand, the matched types correspond to about a third of all the types in the

lexicon-grammar. This can be due to several reasons: an incorrect formalization of the data in the matrix; or some problem in the graphs; or the fact that several proverbs are specific of the Brazilian variant. However, it is interesting to notice that in such large collections, so many types were of the lexicon-grammar were not found. Finally, we remark that only 142 types were common to both collections, corresponding to 347 and 310 matches, respectively. This hints at a large non-overlap of the two lists, not due to variation alone, but mostly to lexical coverage.

The most common type of variation in matched proverbs is the left and right extensions of the core elements: while the keywords of the proverb are found, the graphs also capture variants that present some extra material. This is the case, for example with proverb *Não há regra sem exceção* ‘The is no rule without exception’, from the class P1F1, that produced three different matches (extra material in bold):

Não há regra sem exceção

Em toda a afirmação não há regra sem exceção

Não há regra sem exceção, ***nem mulher sem senão***

As one can see, the first line corresponds to the basic form of the proverb, while in the following lines there is a fronted preposition phrase and coordinated clause, respectively.

Variation in vocabulary is also observed. The most affected grammatical classes are nouns and verbs, showing morphological and lexical variation but no differences regarding the proverbs meaning or function. The next two proverbs illustrate these types of variation, respectively (keywords in bold):

O ***fim*** ***justifica*** os ***meios***

Nem sempre os ***fins*** ***justificam*** os ***meios***

Os ***fins*** não ***justificam*** os ***meios***

Burro velho não ***aprende letra***

Burro velho não ***aprende línguas***

Burro velho não ***aprende o caminho***

Representing in the lexicon-grammar most keywords by their lemma already captures morphologic variation. However, in order to capture the lexical variation, since not all variants are yet represented in the lexicon-grammar, a good strategy would be to semi-automatically create new graphs replacing one keyword at a time by its PoS code (e.g. <N> for the nouns, <V> for the verbs, and so on), and then automatically retrieve the variants that were not yet registered in the lexicon-grammar.

4. CONCLUSION AND FUTURE WORK

This paper presented a general framework for the formal (syntactic) classification of proverbs, based on their structure and syntactic properties. The first steps towards the construction of a lexicon-grammar of Portuguese proverbs by building a database of proverb types, consisting of the keywords that univocally identify each proverb, and allowing for some lexical and morphologic variation. Preliminary experiments on the conversion of these lexical matrices into finite-state tools were carried out, which enable to identify, delimit and tag proverbs and their variants in texts. An extensive list of 56,150 proverbs (and variants), taken from two large collections was digitized and will function

not only as a source for the completion of the lexicon-grammar, but also as a testing ground for the finite-state tools built to identify proverbs in texts.

From the obvious contrast between the number of proverbs from each collection and the number of types already represented in the lexicon-grammar, the next first step is to extend the lexical coverage of the matrix. One of the methods to be explored is the semi-automatic construction of new graphs replacing each content keyword (mainly nouns, verbs and adjectives) by the corresponding PosS. As a consequence, the initial classification may need to be refined, particularly in the case of the most productive structures, which may require sub-classification or even the creation of new classes. Attention must be given to devise a reference procedure that will allow relating, in a single lexical-grammatical unit, those variants belonging to different formal classes, often involving changes in word order. A more precise understanding of the variation phenomena, and the construction of reference graphs for each formal class will reflect of a more precise definition of the insertion window and the sentence alternations that proverbs often allow. It is also a future objective, once a satisfactory lexical coverage is achieved, to measure the frequency of the proverbs types (and their variants) in real corpora, in order to associate frequency information to the database.

Acknowledgements

Research for this paper was partially funded by Portuguese national funds through Fundação para a Ciência e a Tecnologia, under project PEst-OE/EEI/LA0021/2014.

References

- ALMEIDA, J.J. 2014. *Dicionário aberto de calão e expressões idiomáticas*. [online] Available at: <<http://natura.di.uminho.pt/~jj/pln/calao/dicionario.pdf>> [Accessed 14 March 2015].
- BRUCKSCHEIN, M., MUNIZ, F., SOUZA, J., FUCHS, J.T., INFANTE, K., GONÇALEZ, P.N., VIEIRA, R. and ALUISIO, S.M. 2008. *Anotação linguística em XML do corpus PLN-BR. Série de Relatórios do NILC*, São Carlos (SP): NILC-ICMC-USP.
- COSTA, J. R. M. 1999. *O Livro dos Provérbios Portugueses*, Lisboa: Editorial Presença.
- MACHADO, J.P., 2011. *O Grande Livro dos Provérbios*, 4ª ed., Lisboa: Casa das Letras.
- RASSI, A.P., 2014. *List of Proverbs in Brazilian Portuguese*. <https://www.researchgate.net/publication/269165152_Rassi2014> [Accessed 14 March 2015]. DOI: 10.13140/2.1.4907.7280
- RASSI, A.P., BAPTISTA, J. AND VALE, O. 2014. Automatic Detection of Proverbs and their Variants. In: M. Pereira, J. Leal, J. and A. Simões, eds. *Proceedings of the Symposium on Languages, Applications and Technologies (SLATE'14)*. Leibniz (Germany): Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Dagstuhl Publishing. pp. 235-249.
- ROCHA, P. AND SANTOS, D. 2000. CETEMPúblico: Um corpus de grandes dimensões de linguagem jornalística portuguesa. In: Nunes, M.G. et al., eds., *V Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR'2000)*, São Paulo: ICMC/USP. pp. 131–140.

Capítulo 2

Let's play with proverbs?

NLP tools and resources for iCALL
applications around proverbs
for PFL

Artigo 2

Publicado em: *Proceedings of the International Congress on Interdisciplinarity in Social and Human Sciences*, 5th-6th May, pp. 427-446

LET'S PLAY WITH PROVERBS? NLP TOOLS AND RESOURCES FOR ICALL APPLICATIONS AROUND PROVERBS FOR PF

Sónia Reis
Universidade do Algarve
(reis.soniamm@gmail.com)

Jorge Baptista
INESCID Lisboa, Spoken Language Laboratory
(jbaptis@ualg.pt)

ABSTRACT

Proverbs are an important form of cultural expression of a society and are related to various areas of knowledge and human experience (González Rey, 2002). While linguistic elements in widespread use, proverbs are very rich structures both from a cultural and from a linguistic point of view and can therefore contribute significantly to the teaching of languages, both native and foreign (Council of Europe, 2001). However, though there are extensive collections of Portuguese proverbs with tens of thousands of forms and its variants (Reis, *in preparation*), its automatic identification in texts is quite difficult, given its formal variation, both lexical and syntactic (Chacoto, 1994). Nevertheless, using real examples, where proverbs are used in a natural or spontaneous discourse context, is a more natural way to learn and teach the complex conditions and communicative situations that determine the use and meaning of these expressions. On the other hand, frequency indices associated with proverbs and its variants would allow one to select the most common expressions. These are precisely the most interesting forms from the point of view of their teaching/learning and could serve as a basis for the construction of educational games, particularly for learning Portuguese autonomously as a foreign language (PFL) assisted by computer. To make this possible, it is necessary, first of all, be able to recognize the occurrence of proverbs in the texts (Rassi *et al.* 2014), including the instances where these expressions are presented in a truncated or creatively modified form, for example, to better suit the communicative situation or to produce new and more expressive meanings. In this paper, we present an ongoing project, which aims at automatic identification of proverbs in texts. In this interdisciplinary study, we combine natural language processing tools with questionnaires construction techniques for teaching purposes (Hoshino and Nakagawa 2005, Correia *et al.* 2010). This is illustrated here with different sets of formats that can be built based on the knowledge of the form and variation of proverbs, as well as their frequency in *corpora*.

Keywords: Portuguese Proverbs, Intelligent Computer Assisted Language Learning (iCALL), Natural Language Processing (NLP), Didactic Gaming.

JEL Classification: Z00.

1. INTRODUCTION

Proverbs are an important form of cultural expression of a society and are related to various areas of knowledge and human experience (González Rey, 2002). This type of expressions has a relatively frozen structure, usually formed by short phrases, and often composed of two (sometimes three, or more) parts. Also, it is common to find alliterations and rhymes in proverbs. These are all mnemonic features denoting their transmission process, which is predominantly oral, and they generic value as words of wisdom (maxim, precept, etc.). In spite of its frozen characteristics, because of their oral tradition, they also present a relatively wide spectrum of lexical and syntactic variation (Chacoto, 1994). This formal variation makes it particularly difficult the automatic identification of proverbs in texts, which constitutes a challenge to Natural Language Processing (NLP). However, their automatic detection and delimitation not only would be relevant for natural language understanding and automatic discourse analysis, as it could also prove to be extremely useful for several NLP applications (Witten *et al.* 2011), such as machine translation, or even to language learning and teaching. It is on this latter field of application that this paper will focus.

In order to use proverbs as linguistic material for language learning and teaching, it is necessary, in the first place,

to be able to retrieve them from the texts where they occur, and especially those that occur more frequently and in a larger set of communicative contexts. In order to be able to produce such careful selection for pedagogic purposes, it is important that frequency indices be associated to the proverbs and their variants.

In this paper, we focus on the types of activities (didactic exercises) that it is possible to develop around proverbs in the framework of computerassisted language learning (CALL). More precisely, we adopt the CALL perspective inspiring the REAP.PT platform, presenting a brief overview of its most salient features, in order to draw a set of exercises highlighting the linguistic resources and NLP tools required to build such complex objects. Our main goal is to set up a roadmap in view of developing those applications, which we expect to be able to build, in the future.

The paper is organized as follows: In 1.1. some general issues about proverbs and their role in language learning are addressed, particularly in an iCALL context for Portuguese as a Foreign Language (PFL). In 2, existing online exercises are presented and commented. In 3, the main NLP tools and resources for NLP helpful for building exercises on proverbs are briefly presented. In 4, the *corpus* of proverbs and its construction is briefly described. In 5, some remarks are made on the automatic identification of proverbs in texts, and in 6 we present several suggestions regarding the automatic generation of exercises with proverbs using the NLP tools and language resources mentioned in the preceding sections. Finally, in 7, the paper is summarized and future work is considered.

1.1 General Issues on teaching proverbs in a iCALLPFL environment

While linguistic elements in widespread use, proverbs are very rich structures both from a cultural and from a linguistic point of view and can therefore contribute significantly to the teaching of languages, both native and foreign. According to the *Common European Framework of Reference for Languages* (Council of Europe, 2001), the communicative language competence comprises three components: (a) linguistic competences; (b) sociolinguistic competences and (c) pragmatic competences. *Lexical competence*, an integral part of linguistic competences, concerns the “knowledge of, and ability to use, the vocabulary of a language” (*idem*: 110) and it includes the knowledge of both grammatical and lexical elements, including proverbs and other fixed expressions. “Sociolinguistic competence is concerned with the knowledge and skills required to deal with the social dimension of language use” (*idem*: 118), and also includes proverbs since they are “fixed formulae which both incorporate and reinforce common attitudes, make a significant contribution to popular culture” (*idem*: 120).

Information technologies have progressively acquired a more important role in teaching, and particularly in language teaching. Computerassisted language learning platforms allow students to have a more active role in the learning process, enabling them to exercise in an autonomous way their skills and, at the same time, providing constructive feedback on their progress. They are also an important tool to monitor students’ performance and progress, giving the teacher the means to target, in a more personalized way, his/her intervention.

The REAP.PT⁷³ is a computerassisted languagelearning platform developed for the teaching and learning of Portuguese. It is a powerful aid to the learning process, both for native speakers and for students of Portuguese as a foreign language, and it also constitutes a valuable auxiliary tool to the teacher. The system is specially aimed at the learning of vocabulary and it is based on the intensive use of Natural Language Processing (NLP) tools and techniques in order to build automatically a large variety of exercises, drawn from *corpora* of real texts, with their contents adapted to the topic preferences of the students. Many exercises consist of multiplechoice questions, in which a target expression (in the question) is presented (in the set of possible answers) along with a set of *distractors* (a.k.a. *foils*), usually no more than three, each one selected based on a set of linguistically motivated criteria. In this process, the sentences with the target expression are automatically retrieved from *corpora*, while the set of distractors is automatically built using NLP tools and techniques, using with linguistic resources adapted to that purpose. Furthermore, the REAP.PT system is conceived in order to be able to build a dynamic model of the student, keeping track of his/her progress and allowing the teacher to monitor it. Finally, the system, and most prominently, the system automatically correct the exercises and, as much as possible, produces some positive feedback about the linguistic aspect at stake, or alerting the student for the nature of incorrect answer (Pellegrini *et al.* 2012).

This paper intends to lay down the basis for the construction of didactic games with proverbs, in view of integrating them into the REAP.PT platform. With this games involving proverbs, students will be able to identify and use this type of culturally rich trove of expressions, hone their language skills, and evolve their vocabulary competence. To this end, we will first present an overview of the type of exercises involving proverbs that can be found in different websites and computerassisted language learning platforms, highlighting the challenges each type poses to its construction under the REAP.PT approach to language gamming, focusing on the type of linguistic resources and natural language processing techniques required to put them into practice.

⁷³ <http://www.l2f.inescid.pt/wiki/index.php/REAP.PT> (last access: March 31, 2016; all the remaining URL mentioned in this paper have been check on this date).

2. SURVEY OF LANGUAGE RELATED EXERCISES WITH PROVERBS

In this section, we present a brief survey of existing exercises built around proverbs or focusing on the learning form and meaning, including certain CALL platforms, especially in view of learning Portuguese as a foreign language.

A large part of the games and exercises on proverbs made available in several websites adopt a format whose goal is to complete the expression, either half of the proverb is provided so that the user is supposed to know and complete the other half, or one or more words had been removed and the uses had to fill in the corresponding blank spaces. The purpose of these games and exercises, besides promoting the better knowledge of the proverbial stock of the language, can sometimes be inferred as targeting several language skills. In the following lines, we present some of the most interesting of these websites' games and exercises.

The game *Finish the Proverb*⁷⁴ consists, basically, in a multiplechoice answering task, completing the proverbs from which one or more words have been removed. Several exercises (10) are presented at a time, each featuring a proverb that must be completed, and a set of four answers is provided, one being the correct answer and the other three the distractors (or foils). The exercise is automatically corrected but only after the entire set of questions has been answered (and not one at a time). In the correction, the user gets to see the correct answer for each question. The questions form a closed set and, as far as we could find, they have been manually constructed. One of the interesting aspects of this site is the fact that, in the feedback, the user is shown the percentage of players that got that answer right (but not the absolute number!). This could be an indicator of the frequency of the lexical availability of those proverbs. The meaning of each proverb is also explained.

The set of proverbs produced in this game aim at different linguistic competences from the player. In the next table, we try, in a nonexhaustive way, to identify those language skills. These linguistic aspects concern both the target word and the distractors, as well as the interplay between them, so that, on one hand, their decoding is not always easy, and on the other hand, it is sometimes difficult to grasp exactly the concepts at stake, often more than once at the same time. Hence, these are complex exercises, plainly justifying the human effort put into their construction, and highlighting the challenge that their automation involves. Table 1 provides a detailed analysis of the type of relations the distractors hold with the target word, for each proverb.

Several linguistics phenomena are dealt with in these proverbs, as the comments provided for each distractor and highlight, which are associated either to the structure or to the meaning of these expressions. In other words, distractors are never random, each has been produced with a purpose in mind. Each aspect that a distractor involves provides the student with the opportunity to test his/her linguistic competences, and often several of them at the same time, particularly his/her lexical and semantic knowledge. All this is done in a nonexplicit way but still very much "applied" situation, while the gaming context, with the scoring or agonistic situation, entails a more appealing stimulus. Among the relevant phenomena, the most prominent are: (i) the semantic relations between word senses (synonymy and antonymy; holonymy and meronymy; hypernymy and hyponymy; causeeffect) (ii) symbolic values invested in lexical items, having to do with culture and worldknowledge; (iii) collocational nature of word combinations⁷⁵; and (iv) formal aspects such as rhyme, the two or threefold partition of proverbs (*hemistichs*), and paronymy.

Table 1: Linguistic categories/phenomena and language skills mobilized in the task of completing proverbs

Question	Proverb & Answers	Linguistic phenomena
1	Hatred is as blind as ... (D1) A mole (CA) Love (D2) Darkness (D3) None of these	Correct answer: antonymy <i>hatred/love</i> and a semantic relation with the proverb <i>Love is blind</i> <i>Distractor 1:</i> relation with the member of a set (animal/mamífer) and analogy with the frozen idiomatic adverb <i>to blind as a bat</i> <i>Distractor 2:</i> semantic association between <i>blindness</i> and <i>darkness</i> <i>Distractor 3:</i> something else (unknown target)
2	Blood is thicker than ... (D1) Wine (CA) Water (D2) Tears (D3) Sweat	Correct answer: (unmotivated, idiomatic) <i>Distractor 1:</i> semantic but culturallyfounded relation (in Christian tradition, the eucharistic wine is associated with blood) <i>Distractor 2:</i> member of the set of objects to which <i>blood</i> can be associated, and which could be subsumed under the term <i>bodily fluids</i> <i>Distractor 3:</i> same as above

⁷⁴ www.123facts.com/playquiz/FinishTheProverb5244.html

⁷⁵ The term collocation is used here in the sense of Meřčuk & Polguère (2007): a word (the *collocate*) that establishes a well defined semantic relation (*lexical function*) with another word (the *base*); the choice of the collocate is lexically determined, usually idiosyncratic. Lexical functions are a finite, though large, set of semantic and syntactic relations such as qualitativepositive assessment BONUS ("good"), which is the case here: given the base *soil* a finite set of adjectives exist to express its quality (in view of its use) *rich*, *nutritious*, etc.

Question	Proverb & Answers	Linguistic phenomena
3	A stitch in time saves ... (D1) Thousands (D2) A dime (CA) Nine (D3) Everybody	Correct answer: rhyme <i>time/nine</i> <i>Distractor 1:</i> both <i>a</i> and the numeral <i>thousands</i> are determiners <i>stitch</i> ; the numeral is hyperbolic (= <i>many stitches</i>) and it is the opposite of <i>a</i> (single) <i>stitch</i> <i>Distractor 2:</i> rhyme <i>time/nine</i> <i>Distractor 3:</i> the distractor is an indefinite pronoun distributionally adequate as direct object of the verb, thus producing the generic reading typical of proverbs
4	Tall oaks grow from ... (D1) Strong earth (D2) Nutritious soil (D3) Small saplings (CA) Little acorns	Correct answer: antonymy <i>tall/little</i> and semantic relation between <i>oak</i> and <i>acorn</i> ; <i>Distractor 1:</i> the noun <i>earth</i> is semantically motivated as locative complement of <i>grow from</i> ~ and is quasisynonymous of <i>soil</i> ; <i>strong</i> is a collocational modifier of <i>soil</i> ; <i>Distractor 2:</i> <i>idem</i> ; <i>nutritious</i> is also a collocational modifier of <i>soil</i> ; <i>Distractor 3:</i> antonymy <i>tall/small</i> and semantic relation between <i>oak</i> a (tree species) e <i>sapling</i> (a development stage of a tree).
5	The ... wounds more than a lance (D1) Hand (CA) Tongue (D2) Word (D3) Evil eye	Correct answer: (unmotivated, idiomatic): the bodypart <i>tongue</i> connotes with <i>language</i> ; <i>Distractor 1:</i> metonymy between <i>hand</i> and the agent (human) subject of <i>wound</i> ; contrast between agent and the instrument (<i>lance</i>) <i>Distractor 2:</i> <i>word</i> also connotes <i>language</i> ; <i>Distractor 3:</i> the compound <i>evil eye</i> is formed with a bodypart noun (like CA <i>tongue</i> and D1 <i>hand</i>); contrast between a physical instrument (<i>lance</i>) and a symbolic one (<i>evil eye</i>).

In this same website, there are other quizzes involving proverbs. *Finish the Proverb Vol.2*⁷⁶ is quite similar to the game presented above, but it also includes some questions whose goal is to identify the meaning of the proverb from four possible answers. This is also the goal of the quiz *Guess the Proverb*⁷⁷, though this also includes questions aimed at identifying the sentence expressing the opposite meaning of the target proverb. Finally, in *Complete the Proverb/Idiom*⁷⁸, is similar to the exercise we presented at length above: ten wellknown idioms and proverbs are produced with a missing word that the player must complete from a set of four possible answers. However, in this case, the feedback indicates that the players guessed 100% of answers correctly, which may indicate that this quiz is too easy, and may not be challenging enough for the user. To conclude, it is worth mentioning that the site allows players to rate, bookmark and recommend it to other users.

Table 1 (cont.)

Question	Proverb & Answers	Linguistic phenomena
6	Two wrongs don't make ... (CA) A right (D1) It easier (D2) A fight (D3) The world a better place	Correct answer: antonymy <i>wrong/right</i> <i>Distractor 1:</i> though the sentence is acceptable, the relation with the target word is unclear <i>Distractor 2:</i> Rima <i>fight / right</i> (CA) <i>Distractor 3:</i> though the sentence is acceptable, the relation with the target word is unclear
7	Make a silk purse out of a ... (D1) Silk dress (D2) Woolen sock (CA) Sow's ear (D3) Cow's ear	Correct answer: (unmotivated, idiomatic): opposition between a noble/expensive raw material (<i>silk</i>) and a cheap/gross one, with a negative connotation (<i>sow's ear</i>); <i>Distractor 1:</i> the same material adjective (<i>silk</i>) and a clothing item (<i>dress</i>), which can be associated with <i>purse</i> ; <i>Distractor 2:</i> another material adjective (<i>woolen</i>) outro adjetivo de matéria (<i>woolen</i>) and a clothing item (<i>sock</i>) <i>Distractor 3:</i> paronymy (<i>sow/cow</i>) with the target word, both related to <i>livestock</i> (hypernym); the same bodypart noun (<i>ear</i>)
8	A loaded ... makes no noise (D1) Horse (D2) Gun (D3) Train (CA) Wagon	Correct answer: collocation between <i>loaded</i> and <i>wagon</i> ; appeal to world knowledge, equating <i>make noise</i> with <i>speak loudly</i> and <i>loaded</i> with <i>richness</i> ; <i>Distractor 1:</i> collocation between <i>loaded</i> and <i>horse</i> ; <i>Distractor 2:</i> collocation between <i>loaded</i> and <i>gun</i> ; <i>Distractor 3:</i> collocation between <i>loaded</i> and <i>train</i> ; meronymy relation between <i>wagon</i> and <i>train</i> ;
9	A burnt child dreads the ... (D1) Match (D2) Stove (D3) Air conditioner (CA) Fire	Correct answer: semantic relation (causeeffect) between <i>fire</i> and <i>burn/burnt</i> ; <i>Distractor 1:</i> semantic relation between <i>fire</i> and <i>match</i> (instrument); <i>Distractor 2:</i> semantic relation between <i>fire</i> and <i>oven</i> (object/tool to generate heat) and, indirectly, a semantic relation (causeeffect) between <i>oven/heat</i> and <i>burn/burnt</i> ; <i>Distractor 3:</i> semantic relation between <i>air conditioner</i> and the distractor <i>stove</i> (both subsumed under the hypernym <i>domestic appliances</i>), also involving the concept of <i>temperature</i> .

⁷⁶ <http://www.123facts.com/playquiz/FinishTheProverbVol25305.html>

⁷⁷ <http://www.123facts.com/playquiz/GuessTheProverb5297.html>

⁷⁸ <http://www.123facts.com/playquiz/CompletheProverbIdiom7046.html>

Question	Proverb & Answers	Linguistic phenomena
10	Learn to walk before you ... (D1) Crawl (D2) Speak (CA) Run (D3) Leap	Correct answer: <i>walk</i> and <i>run</i> are both motion verbs but <i>run</i> implies more speed, thus there is a gradation that motivates the order of these verbs in the proverb; <i>Distractor 1:</i> the relation between <i>crawl</i> and <i>walk</i> is similar as in <i>walk/run</i> (CA) but the gradation is inverted; <i>Distractor 2:</i> <i>walk</i> and <i>speak</i> represent two successive stages in child development; <i>Distractor 3:</i> the relation between <i>leap</i> and <i>walk</i> is similar (motion verbs) but the comparing parameter is not <i>speed</i> but rather <i>complexity</i> or <i>distance</i> .

CA= Correct answer; D1= Distractor 1; D2= Distractor 2; D3= Distractor 3; the order of the answers, which may also be relevant for the task, is the same as shown in the website. Apparently, this order is constant.

The *Centro Virtual Camões*⁷⁹ is one of the Portuguese websites that produces a set of language-related games/quizzes. The *Jogo da Lusofonia*⁸⁰ is a game about the Portuguese-speaking countries, providing 256 multiple-choice questions, drawn from different topics, including some proverbs and idioms. The player, who can choose to play 5, 10 or 15 questions at a time, can customize each run. The user also has 3 opportunities to ask for help. Answers consist of 3 choices and are timed (30 seconds each), which adds to the interest of the game. The game can be played alone or with a team, and with a number of other playing partners (or teams) as adversaries. The goal of the game is to achieve the maximal number of correct answers in the least time.

As far as the proverbs are concerned, most questions in this game consist in completing the proverb or choosing its correct meaning. The questions have been manually built and the authors have also produced the correction. In the feedback, along with the correct answer, the meaning of the proverb is also explained. Due to the overall goal of the game, some of the proverbs found in the game may not be part of the common stock of Portuguese-speaking countries (at least they were absent from the (European) Portuguese collections of proverbs known to us). For example, the Guinean proverb: *Os dentes do elefante não o cansam* ‘The teeth of the elephant do not tire it’ is explained as follows (our translation): ‘In Guinea-Bissau, the elephant is a symbol for parents, who, in spite of the material difficulties, are tireless in caring and raising their children’. However, in our European culture, elephants do not have the same connotations and shown by the five proverbs⁸¹ we find in our *corpus*, where this animal is mentioned:

- a) *A sebe dura três anos, o cão três sebes, o cavalo três cães, o homem três cavalos, o corvo três homens e o elefante três corvos* (The hedge lasts three years, the dog three hedges, the horse three dogs, three horses a man, the crow three men and the elephant three crows)
- b) *Quando dois elefantes lutam a erva fica espezinhada* (When two elephants fight the grass gets trampled.)
- c) *Com um cabelo de mulher, pode amarrarse um elefante* (With a woman’s string of hair one can tie up an elephant)
- d) *Elefante grandorro de pequerruchinho cresce* (A big elephant grows from a little one)
- e) *O teu inimigo é pequeno como uma formiga, mas guardate dele como se fosse um elefante* (Your enemy is small as an ant, but beware yourself from him as if he were an elephant)

In these proverbs, the elephant can connote with: longevity (a); unwanted/collateral effect of an action, due to the size (b); the overpowering feminine influence over men: contrasting size/strength of a string of hair and an elephant (c); the size/strength of an adult animal can not be guessed from its size as a youngling: contrast between the size of a young and a grown elephant (d); prudence towards an enemy irrespective of its size: contrast between the size of an ant and an elephant.

Besides these proverbs, we also found in the game the idiom *apanhar pulgas e deixar passar elefantes* (lit: catch fleas and let elephants pass, ‘to be mindful of small details but ignore the larger mistakes), which we do not treat as proverb because of its subject being a distributionally free syntactic slot.

In the *Jogo da Lusofonia*, it was not possible to determine how many questions were there about proverbs; still, whenever we played it, we found several questions about proverbs and idioms. Another feature of the game is the feature of presenting on the topleft side the national flag of the Portuguese-speaking country the question is about. Notice, however, that many proverbs from the Brazilian Portuguese also belong to the European Portuguese stock, even if such fact is not signaled by the game. It was not possible to ascertain why was such or such proverb assigned to one the varieties, instead of the other one. Nevertheless, when the correct answer is explain, one often finds the indication that there is a corresponding proverb in more than one of the Portuguese-speaking countries. Fig. 1 is an

⁷⁹ <http://cvc.institutocamoes.pt/>

⁸⁰ <http://cvc.institutocamoes.pt/aprenderportugues/abrir/jogodalusofonia.html>

⁸¹ In the *corpus* there are 16 instances of these proverbs with *elefante* ‘elephant’, distributed by several variants and repeated versions, that is, occurring simultaneously in more than one collection.

example of such type of indication. In this case, the proverb appearing in the question is exclusive of the Brazilian variety of Portuguese, which is hinted by the name of the animal, the *urubu*, a vulturelike species that is endogenous of South America. Furthermore, in this case, the proverb is not exactly explained; instead, a Portuguese equivalent (and yet another, related or almost equivalent) proverb is provided.

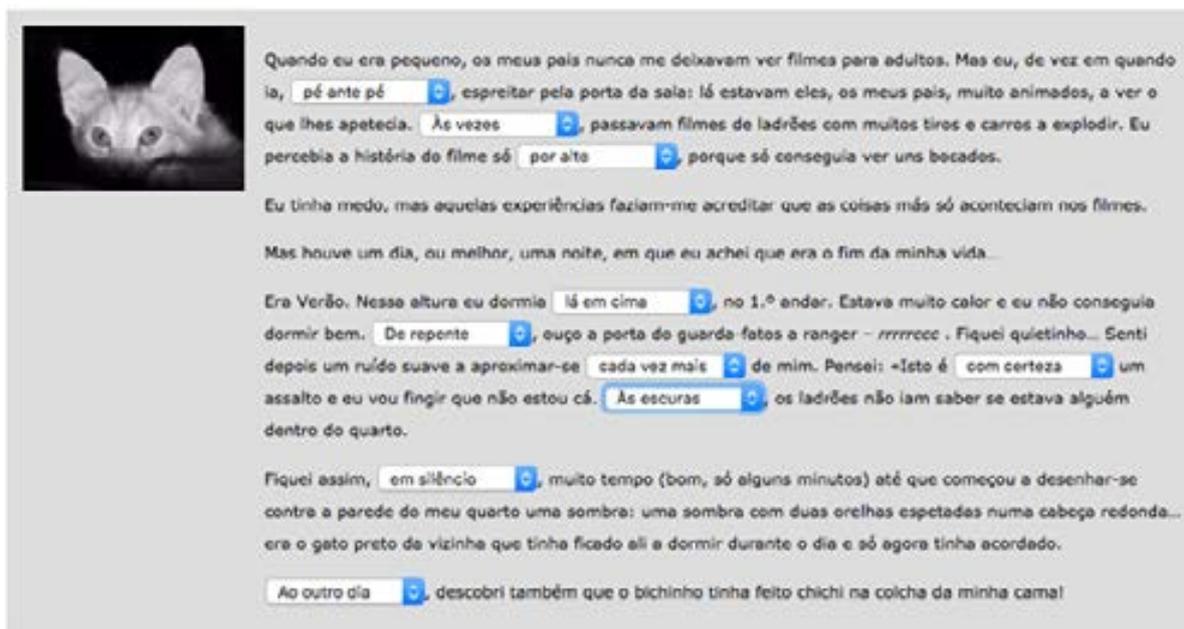
The *Ciberescola da Língua Portuguesa*⁸² is language learning platform that provides support for Portuguese as a Foreign Language (PFL) by making available several interactive language resources and online courses, aimed both at teachers and students, since they can be used within the classroom context, in a collaborative process, as well as a selfstudy tool. This platform covers all level of study, from the 5th to the 12th grade, and it also includes training material for the A2 and B2 levels of PFL. Students can choose their topics of interest (music and other arts, sport, traveling, folklore, etc.) and practice Portuguese in a funny and autonomous way, with exercises aimed at different language competences, namely, the training of oral and written understanding. To the best of our knowledge, no games or exercises involving proverbs have been produced yet for this platform.

Fig. 1: Jogo da Lusofonia



⁸² <http://www.ciberescola.com/>

Fig. 2: Exercise on frozen expressions (“expressões fixas”)



Quando eu era pequeno, os meus pais nunca me deixavam ver filmes para adultos. Mas eu, de vez em quando ia, , espreitar pela porta da sala: lá estavam eles, os meus pais, muito animados, a ver o que lhes apetecia. Às vezes , passavam filmes de ladrões com muitos tiros e carros a explodir. Eu percebia a história do filme só , porque só conseguia ver uns bocados.

Eu tinha medo, mas aquelas experiências faziam-me acreditar que as coisas más só aconteciam nos filmes.

Mas houve um dia, ou melhor, uma noite, em que eu achei que era o fim da minha vida...

Era Verão. Nessa altura eu dormia , no 1.º andar. Estava muito calor e eu não conseguia dormir bem. De repente , ouço a porta do guarda-fatos a ranger – rrrroccc. Fiquei quietinho... Senti depois um ruído suave a aproximar-se , de mim. Pensei: -Isto é , um assalto e eu vou fingir que não estou cá. , os ladrões não iam saber se estava alguém dentro do quarto.

Fiquei assim, , muito tempo (bom, só alguns minutos) até que começou a desenhar-se contra a parede do meu quarto uma sombra: uma sombra com duas orelhas espetadas numa cabeça redonda... era o gato preto de vizinha que tinha ficado ali a dormir durante o dia e só agora tinha acordado.

, descobri também que o bichinho tinha feito chichi na colcha da minha cama!

Concerning idiomatic expressions, only a fillintheblank exercise with frozen expressions (and collocations) has been found (Fig 2). The goal of this exercise, marked as “difficult”, is to fillin the ten blank spaces, and for each one 10 solutions are presented. After all the spaces have been completed, the form is submitted and the player’s score is provided: this is the only feedback the application produces, the number of correct/incorrect answers, but not the solution to the exercise. In some cases, the answerchoices were shown in uppercase, thus hinting on the correct answer whenever that slot was in the beginning of a sentence. Most of these idioms are compound/frozen adverbs (Gross, 1996; Ranchhod, 2003; Palma, 2009): *com certeza* (= *certamente*), *em silêncio* (= *silenciosamente*).

In this exercise, those adverbs express the circumstances of *time* (frequency: *às vezes* ‘sometimes’, daterelative: *ao outro dia* ‘the next day’), *manner* (most of them: *pé ante pé* ‘on tiptoes’, *por aita* ‘vaguely’, etc.), *place* (*lá em cima* ‘upstairs’) and *attitude* (Molinier and Levrier, 2000; *com certeza* ‘certainly’). One of the issues with the exercise comes from the fact that for certain slots more than one possible solution if produced. For example: *Fiquei assim, em silêncio/lá em cima/às escuras, muito tempo* ‘I stayed like that, in silence/upstairs/in the dark’; *Nessa altura eu dormia lá em cima/em silêncio, no 1º andar* ‘At that time, I used to sleep upstairs/in silence’.

The *Observatório da Língua Portuguesa*⁸³ is an association whose main goal is to promote the Portuguese Language. Its webpage presents several resources to help Portuguese learners, including a weekly podcast⁸⁴ for teaching proverbs and idioms to students from the B2 and C2 levels. To date, only 39 idiomatic expressions and proverbs of European Portuguese, a insignificant number in view of the size of the lexicon, are explained with some degree of detail, along with an mp3 downloadable audio file and the corresponding text. In some cases, different expressions are compared, for example, to clarify the distinction between the use of the proverbs *Filho de peixe sabe nadar* ‘The son of a fish knows how to swim’ and *Tal pai, tal filho* ‘Like father, like son’, or relating them with idioms like *ser a cara chapada de alguém* ‘be the spitting image of smb’ or *seguir as pegadas de alguém* ‘follow the footsteps of smb.’ Examples of the use of proverbs are also produced, including some cases of creative adaptation of a proverb to a situation, like the use of feminine equivalent forms in *Tal mãe, tal filha*, to adjust the proverb to female characters. Notice, however, that all these are fabricated examples, and not spontaneous utterances retrieved from *corpora*. This website does not feature any game/exercise involving proverbs, so its usefulness is somewhat limited.

3. LANGUAGE RESOURCES AND NLP TOOLS FOR SEMANTIC PROCESSING

In this section, we present several language resources and NLP tools that are already available for processing Portuguese texts, which may be useful to produce didactic games for learning the proverbs of the language. These resources consist, basically, of a vocabulary where words are explicitly related by a set of semantic relations. For lack

⁸³ <http://observatoriolp.sapo.pt/pt>

⁸⁴ <http://sayitiinportuguese.pt/podcasts/>

of a more consensual term, we call them all *ontologies*.

The *WordNet.PT*⁸⁵ (Marrafa, 2001) is a database developed at the Centro de Linguística of Universidade de Lisboa in the same methodological and conceptual framework as the English WordNet (Fellbaum, 1998)⁸⁶. It is a language resource containing 19,000 expressions from European Portuguese, from different parts of speech and several semantic domains, structured in semantically homogeneous sets (or *synsets*) and the semantic relations between them, such as meronymy (or part-whole relations), synonymy, antonymy, class, event participants and structure. This data can be used in several areas of Computational Linguistics and Language Engineering. More recently, the Brazilian counterpart, *WordNet.Br*⁸⁷ (DiasdaSilva, 2013), includes 5,860 verbs structured in 3,713 synsets.

The *Linguateca* project website⁸⁸ also distributes several semantic language resources. *PAPEL*⁸⁹ (Oliveira *et al.* 2008) is a freely downloadable lexical resource that consists of 102,000 different words and 191,000 semantic relations. These were semiautomatically extracted from the *Dicionário da Língua Portuguesa*, edited by Porto Editora, and include 83,000 (approx. 44%) synonymy word pairs, and (approx. 26%) hyponymy relations.

Table 2: Semantically related entries for the keyword *pássaro* ‘bird’.

Proverb: *Mais vale um pássaro na mão que dois a voar* ‘Better a bird in the hand than two flying’

Language Resource	<i>Pássaro</i> ‘bird’ [semantic relation]: word(lemma)+
WordNet.PT	[quasisynonym]: <i>piupiu</i> [agent/causeresult]: <i>chilreio</i> [corelated with]: <i>áugure</i> [is a member of]: <i>bando</i> [hyponym (type of)]: <i>ave</i> [hypernym (supertype of)]: <i>andorinha, arara, beijaflor, bicodelacre, canário, catatua, corvo, cotovia, cuco, estorninho, gralha, melro, papagaio, pardal, pega, periquito, picapau, pintassilgo, pombo, rola, rouxinol, tarambola, tentilhão, tordo, tucano</i>
PAPEL ⁹⁰	[synonym_N_of]: <i>coucou, melriacho, pássarocoucou, picarei</i> [hypernym]: <i>barrete, lavandisca, pardal, petinha, poupa, ...</i> (172 entries) [is a member of]: <i>ordem_de_pássaro</i> [made_with]: <i>alcapão, aramenha, arapuca, armação, boiz, ...</i> (14 entries)
Onto.PT	[synonym]: <i>ave, passarinho, folecha</i>
MWN.PT	[synonym]: <i>ave</i> [is a member of]: <i>bando, ave</i> [hiponym (is a kind of)]: <i>vertebrado</i>
PULO	[synonym] ⁹¹ : <i>ave</i>
TEP 2.0	not found

*Onto.PT*⁹² (Oliveira & Gomes, 2014) aims at building a lexical ontology for Portuguese, with information extracted from six different resources (dictionaries, thesauri, encyclopedias as well as from other ontologies and *corpora*⁹³). It is also a freely downloadable ontology, and contains 156 lexical forms, organized in 117,000 *synsets* and more than 173,000 tuples and their semantic relations.

The *MNW.PT*⁹⁴ MultiwordNet of Portuguese (Pianta *et al.* 2002) covers manually allegedly validated 17,200 synsets, connected by hyponymy and hyperonymy relations, concerning over 21,000 word senses/forms and 16,000 lemmas of European and Brazilian Portuguese. It also includes subontologies for the categories Person, Organization, Event, Location and ArtWorks, and 98 core base concepts, suggested by the Global Wordnet Association, along with 164 concepts defined by the EuroWordNet project.

*PULO*⁹⁵ (Simões & Guinovart, 2014) is a Portuguese Unified Lexical Ontology and works as a wordnet compatible

⁸⁵ <http://www.clul.ul.pt/clg/wordnetpt/index.html>

⁸⁶ <http://wordnet.princeton.edu/>

⁸⁷ <https://en.wikipedia.org/wiki/WordNet>

⁸⁸ <http://www.linguateca.pt>

⁸⁹ www.linguateca.pt/PAPEL

⁹⁰ There were in total 220 words associated to *pássaro* ‘bird’. Due to this high number, only some few entries are shown here.

⁹¹ The only information available is that the word belongs to the same synset. For consistency, we treated as a synonym.

⁹² <http://ontopt.dei.uc.pt/>

⁹³ Namely the *Dicionário PRO da Língua Portuguesa 2005*, da Porto Editora, through the PAPEL project, the *Dicionário Aberto* (dicionarioaberto.net), Wikcionário.PT (<https://pt.wiktionary.org>), TeP 2.0 (<http://www.nilc.icmc.usp.br/tep2/>), the *OpenWordNetPT* (<https://github.com/ownpt/openWordnetPT>) and the *OpenThesaurus.PT* (<http://paginas.fe.up.pt/~arocha/AED1/0607/trabalhos/thesaurus.txt>).

⁹⁴ <http://mwnpt.di.fc.ul.pt/>

⁹⁵ <http://wordnet.pt/>

with wordnets available for other languages (English, Galician, Basque, Spanish and Catalan), using Probabilistic Translation Dictionaries automatically created from parallel *corpora*. The process of bootstrapping an European Portuguese WordNet from the English, Spanish and Galician wordnets generated a total of 56,770 synsets and 97,058 variants.

*TeP 2.0*⁹⁶ (Maziero *et al.* 2008) is a vocabulary of Brazilian Portuguese, also developed from the Princeton wordnet framework, where words are grouped together in sets, establishing a synonymy relation (and in some cases antonymy), and providing examples for (some of) the queried words. The four major POS (noun, verb, adjective and adverb) are represented in TeP, which is freely available, and can be downloaded and used for many applications.

Table 3: Semantically related entries for the keyword *amigo* ‘friend’.

Proverb: *Quem te avisa, teu amigo é* ‘He who warns you is your friend’

Language Resource	<i>amigo</i> ‘friend’ [semantic relation]: word(lemma)+
WordNet.PT	[quasiantonym of (noun)]: <i>inimigo</i> [corelated with (noun)]: <i>amizade</i> [hyponym (type of) (noun)]: <i>pessoa</i> [hypernym (supertype of) (noun)]: <i>amigalhão, camarada, confidente</i> [quasiantonym of (adjective)]: <i>inimigo</i>
PAPÉL	[synonym_N_of]: <i>afeiçoado, aliado, amante, camarada, camba, chamar, chapa, companheiro, partidário, simpatizante, xará</i> [synonym_ADJ_of]: <i>afeiçoado, aliado, unido</i> [hyperonym]: <i>amigalhão, antropófilo, copista, demófilo, filósofo, pendenciador, reinadio</i> [is a member of]: <i>peçoal</i> [antonym_ADJ_of]: <i>inimigo</i> [made_with]: <i>traição</i>
Onto.PT	[synonym (adjective)]: <i>afeiçoado, aliado, amante, amical, amigável, amigo, amistoso, apegado, apreciador, camarada, colaço, contubernal, ligado, pegado, unido</i> [synonym (noun)]: <i>achegado, acompanhante, aderente, afeiçoado, aliado amante, amásio, amizade, barregão, camarada, camba, capeba, chamar, chapa, colega, companheiro, companhom, compincha, concubino, confrade, conhecido, contubernal, dedicado, inclinado, irmão, malungo, miça, mirmidão, partidário, simpatizante, xará</i>
MWN.PT	[hyponym of] <i>criatura, indivíduo, pessoa, ser humano</i> [is a kind of] <i>alter ego, camarada, colega de apartamento, colega de escola, colega de quarto, compadre, companheiro, compincha, cupincha (BR), confidente, irmão</i>
PULO	[synonym] <i>aliado, aficionado, amante, amizade, colega, companheiro, irmão, seguidor, sócio</i>
TEP 2.0	[synonym (adjective)]: <i>afeiçoado, aliado, amical, amigável, amistoso, apegado, camarada, colaço, contubernal, ligado, pegado, unido</i> [synonym (noun)]: <i>afeiçoado, amante, amásio, camarada, companheiro</i>

Since most of the existing gaming exercises with proverbs feature several semantic relations between the distractors and the target word, we tried to assess the usefulness of some of these language resources in view of the automatic generation of didactic exercises, since they explicitly encode semantic relations between words. For two very common proverbs, appearing simultaneously in the 4 collections of proverbs we used to establish our *corpus* (section §4), we retrieved the words (potential distractors) these resources could produce when queried for entries that are in some semantic relation with the proverb keywords (the targets). The results are shown in Tables 2 and 3.

One of the issues that has to be addressed in the automatic generation of distractors for this type of exercise in the gendernumber agreement between, on one hand, the target word (*v.g.*, *pássaro* ‘bird’: masculinesingular) and the surrounding words in the proverbs (*v.g.*, *um* ‘a/one’: masculinesingular, *dois* ‘two’: masculineplural); and, on the other hand, the gendernumber of the distractors (*v.g.*, *ave* ‘bird’: femininesingular). Therefore, when retrieving the distractors, either these gendernumber features act as filters, selecting only those words that match the target word values; or some the surrounding words of the proverb may have to be adjusted (or the distractor inflected appropriately) to keep the sentence grammatically correct. For example, the two proverbs would need to be adapted to produce the following, grammatically correct strings (distractors shown [inside square brackets] and modified word in **bold**):

*Mais vale **uma** [ave] na mão do que **duas** a voar*

‘It is better to have a/one bird in the hand than two flying’

This latter option requires, however, the sentence be previously parsed in order to determine which words feature the gendernumber agreement with the target (basically determiners and modifiers, or the verb if the target word is

⁹⁶ <http://www.nilc.icmc.usp.br/tep2/index.htm>

the head of the subject). This is made more complex when nonlocal constraints are in place, as in this case, where *anaphora resolution* (Mitkov 2002, Marques 2014) is required: the numeral *dois* is an anaphor introducing a repeated (and, therefore, zeroed) instance of the target noun⁹⁷:

Mais vale um pássaro na mão do que dois (pássaros) a voar

In this case, only one masculine noun is present in the sentence, that can be the antecedent of the zeroed word. However, in the case of the distractor *ave* 'bird', when producing the morphosyntactic adjustments for the gender/number agreement, two possible antecedents are available, the distractor proper and the word *mão* 'hand', also a feminine noun.

Mais vale uma ave na mão do que duas (aves) a voar

On top of it all, this inflection information is not available in lexical resources, since their purpose is different. Therefore, it is necessary to associate that information to the data retrieved from the ontologies, both for parsing the proverbs and to generate the adequate inflected forms for the morphosyntactic adjustments. Since there are already several computational lexicons of Portuguese (for example, Bick, 2000; Ranchhod *et al.*, 1999, Vicente, 2013, among others), this should not constitute an insurmountable subtask, though it involves some additional processing.

On the other hand, for the parsing stage, the natural language processing required is much more complex and only some few systems exist, for Portuguese, that could produce such data. Among others, one could cite PALAVRAS (Bick, 2000), LXParser (Silva *et al.* 2010) and STRING (Mamede *et al.*, 2012).

Another issue comes from the fact that some of the items retrieved from the ontologies under the relation of synonymy are not exactly so, probably because some of these resources were automatically derived from existing dictionaries (for human users). For example, the following words were indicated as synonyms of *amigo* 'friend': *conhecido* 'acquaintance', *irmão* 'brother', *sócio* 'partner', *aficionado* 'id.', but only those that can appear a predicative (postcopula) context with a human determinative complement (*X be N of Y*) are adequate to replace the target word. Thus, *aficionado* 'id.' does not have such properties and, therefore, it can not enter the proverb's structure (**Quem te avisa teu [aficionado] é*). On the other hand, for antonyms, one finds *inimigo* 'enemy'. This produces a semantic drift that may engender acceptability issues in the resulting distractors.

Naturally, some resources have limited usefulness since they do not feature the target words, or do not include any items from the same POS as the target. For example, TeP 2.0 has no entry for *pássaro* 'bird', a very common word, and MWN.PT only contains nouns, so it is useless for replacing verbs.

4. CORPUS

Several (printed!) collections of proverbs are already available in the literature, but to the best of our knowledge, only one digital resource exist, with about 2,300 proverbs and (some) variants⁹⁸. To fill in this important gap in available linguistic resources, a *corpus* with +114,000 proverbs and their respective variants has been built (Reis, *in preparation*), digitized from four printed collections of proverbs (Costa, 1999; Machado, 1996; Moreira, 1996; Parente, 2005) and semiautomatically corrected. Each proverb has been identified unambiguously by a conventional code, allowing the user to recover the source it came from.

One of the first tasks to make this new *corpus* useful consisted in determining which proverbs appeared in more than one of the source collections. This aims at identifying the most usual proverbs, since different authors would have collected them more than once and independently. However, achieving that goal is not trivial, since proverbs present a wide variation both orthographic and in the use of punctuation. Other proverbs are clearly variants, only differing slightly, in the use of a determiner, a preposition, or other grammatical words. Still, allowing for this difficulty, 44,272 (approximately 39%) of the proverbs appear more than once in the *corpus*, while 70,139 (approximately 61%) appear only once. Due to those variation factors (and remaining scanning/optical character recognition errors), it is likely that the number of repeated proverbs might increase.

⁹⁷ In fact, this is a special type of anaphora, the so-called *lexical anaphor*, since the zeroed *anaphor* refers to the *lexical item* (the noun) in the antecedent and not to the same extralinguistic entity that word refers to, as in a typical anaphoric relation.

⁹⁸ <http://natura.di.uminho.pt/jjbin/dac>

Table 4: Distribution of the proverbs' keys by frequency classes (bins) in the *corpus*

Frq	Count	%	Frq	Count	%	Frq	Count	%
1	20,453	0,39	8	108	0,00	15	0	0,00
2	12,261	0,24	9	53	0,00	16	0	0,00
3	10,805	0,21	10	22	0,00	17	0	0,00
4	6,184	0,12	11	9	0,00	18	0	0,00
5	1,132	0,02	12	7	0,00	19	1	0,00
6	470	0,01	13	6	0,00	20	0	0,00
7	273	0,01	14	1	0,00	total	51,785	1,00

Since our ultimate goal is to group together variants of proverbs in order to be able to take that set of forms as a single *paremiologic unit*; and, as we intend to do so by associating to each proverb a set of distinctive keyelements, in order to be able to identify proverbs univocally in texts; we proceed with the construction of a **key** to each instance of the *corpus*, by automatically performing a set of operations on the strings of words that make up that large list. These operations included: (a) convert all text to lowercase; (b) remove all punctuation marks; (c) “cleaning” all repeated white spaces and incorrectly split words (due to the OCR); and (d) removing all *stopwords*⁹⁹, after establishing a list of word forms, adequate to the task at hand, from extant lists of stopwords for Portuguese available in the internet. The resulting list of keys contains 51,785 different strings, whose distribution is presented in Table 4.

In spite of its approximate nature, this method allowed, nevertheless, to group together several variants of the same paremiological units, simply by associating them to a key formed of their lexical elements (content words).

5. AUTOMATIC RECOGNITION OF PROVERBS IN TEXTS

To produce certain exercises, it is necessary to be able to retrieve automatically from *corpora* or from the web the texts where the proverbs naturally occur. Recently, Rassi *et al.* (2014) presented a formal (syntactic) classification, based on a *corpus* of 3,500 proverbs and their variants, organized in 595 types (or base forms) and collected from several dictionaries of Brazilian Portuguese proverbs. The authors presented a method to automatically produce finitestate machines using the Unix linguistic development platform (Paumier, 2003, 2016) that can be used to retrieve candidate proverbs from texts. The method was tested in the PLNBr *corpus* (Bruckschen *et al.* 2008), with 29 M tokens of journalistic texts, taken from the online edition of the Folha de São Paulo newspaper (19942005). A precision from 60% to 73% was reported, depending mostly on the formal class of the proverb and the degree of coverage of the variants represented in the list of proverbs. However, the number of proverbs found in this *corpus*, considering the total number of proverbs in the list, is quite small, which probably stems from the journalistic nature of the *corpus*.

Considering the data and the results obtained from this work, namely its reduced coverage of the language stock of proverbs, the issues found in the automatic generation of the finitestate automata, and their impact in the overall precision of the method, another approach has been envisaged. Based on the frequency of the proverbs' keys (produced as described in §), we intend to produce, manually, the finitestate automata (FSA) containing, for each paremiological unit the strings of keywords that uniquely identify it.

To better illustrate the method, let us consider the proverb *Deus escreve direito por linhas tortas* ‘God writes straight with crooked lines’, which is represented in the database by the following entries (the codes on the left represent the source of the variant), in alphabetic order.

LPP_JRMC10132 *Deus escreve por linhas tortas.*

LP_SP11372 *Deus escreve direito (certo) por linhas tortas (travessas).*

GLP_JPM07945 *Deus escreve direito (certo) por linhas tortas.*

PP_AM03569 *Deus escreve direito por linhas tortas.*

GLP_JPM09795 *Escreve Deus às vezes o direito com letras tortas.*

⁹⁹ https://en.wikipedia.org/wiki/Stop_words; Witten *et al.* (2011).

PP_AM04397 *Escreve Deus às vezes o direito por linhas tortas.*

LP_SP14248 *Escreve Deus às vezes o direito por linhas tortas.*

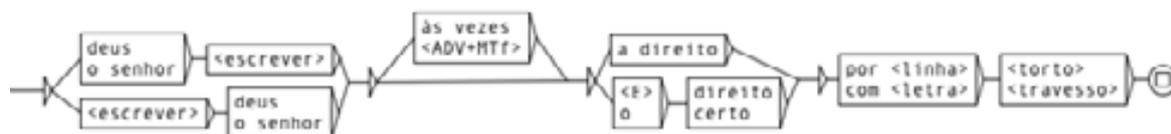
LPP_JRMC10139 *Escreve Deus às vezes o certo por linhas tortas.*

The following variation can be found in these entries can be summarized as follows:

- Subject inversion: *Deus escreve/ escreve Deus* ‘God writes/writes God’;
- Lexical variation: *Deus/ o Senhor* ‘God/the Lord’¹⁰⁰, *direito/ certo* ‘straight’, *linhas/ letras* ‘lines/letters’, *tortas/ travessas* ‘crooked’; and the equivalence of the above with the compound manner adverb *a direito* ‘straightly’¹⁰¹;
- **Nominal use of adjectives** *direto/ certo* ‘straight’, indicated by the article *o* ‘the’;
- (facultative) insertion of frequency adverb *às vezes* ‘sometimes’.

Considering the variation factors above, the following FSA can be built, using the Unitex (Paumier, 2003, 2016) linguistic development platform (Fig. 4)¹⁰²:

Fig. 4: Graph with a finitestate automata representing the formal variation found in proverb *Deus escreve direito por linhas tortas* ‘God writes straight with crooked lines’.



By systematically exploring the graph paths, one can obtain around 120 different strings. A simple query in a search engine of some of these variants is sufficient to retrieve hundreds of instances of this proverb from the web.

Though this is a painstaking process, particularly when there are several, very similar, keys associated with the same proverb, it is possible, in this way, to group together a large number of variants of a paremiological unit, using the large database of collected proverbs described above as source of variation, allowing some degree of free morphologic e syntactic variation (lemmatization, lexical masks, facultative insertions, punctuation), thus aiming at achieving a better performance both in precision and recall in the retrieval of proverbs from texts.

This description work is still at its beginning. Once the coverage of this FSA library is sufficiently large, we expect to use it to retrieve proverbcandidates from texts from real texts, in *corpora* of different genres and text type, as well as from the web. This will allow us to determine the relative frequency of proverbs’ use, drawn from large textual *corpora*, which is an indispensable step towards the construction of didactic applications.

We expect, in this way, to be able to collect a *corpus* of proverbs in their context of use (full texts), thus paving the way to a study on the linguistic devices that introduce proverbs into a discourse, and to model the complex discursive relations they hold with the text around it, linked with the communicative, situationalpragmatic factors surrounding those utterances. Notice that, besides their cultural aspects, this communicative competence is one of the key interests in learning proverbs, both as a native and (which more difficult) as a foreign language.

6. DIDACTIC GAMES: SOME EXPERIMENTS

In this section, we sketch some of the didactic games involving proverbs that can be produced using NLP tools and available resources for Portuguese. Adopting the general framework of REAP.PT, illustrated by the works of (Correia 2010; Pellegrini *et al.* 2012; Correia *et al.* 2012), and briefly outlined above (§), our purpose, here, is just to create a roadmap, defining the games’ goals, and highlighting the complexity of the NLP tasks involved in automatically generating such exercises, eventually drawing the examples from real texts, whenever appropriate.

¹⁰⁰ This variant is not in the *corpus* of collected proverbs, but we found 2 instances in the web (using Google, with exact match and only under the top domain .pt).

¹⁰¹ *Idem*: 6 instances of this variant were also found in the web.

¹⁰² In the graph, the lemma of a word is given inside '<' and '>' (e.g. *escrever*, *linha*, *letra*, *torto* and *travesso*). This allows the system to match any inflected form associated with these lemmas. Notice also that *deus* and *senhor* are written in lowercase to allow case variation, should it appear in texts. The lexical mask <ADV+MTI> represents any adverb of the syntacticsemantic class of temporal (frequency) adverb, both simple and compound (Molinier and Levrier, 2000) Such lexical resources, with the syntacticsemantic information associated with adverbs, have already been built also for Portuguese (Palma 2009, Mamede *et al.* 2012).

Most games already available in the web have been manually produced by specialists (language teachers) in very small and closed batches. These exercises are mostly *multiplechoice* (or *cloze*) *questions*. Each exercise can be seen as a HIT (Human Intelligence Task)¹⁰³ where the student, using a complex interplay of his language competence and world knowledge, has to choose the appropriate word or expression to fill in a blank in the proverb that provided as *prompt*. Besides the correct answer, the student is presented with 3 alternative answers called *distractors* (or foils), to choose from. The selection of the foils is a complex issue, deriving from the objectives set for a specific exercise. In the case of language learning, these objectives are related to the training of linguistic competences, mostly of semantic nature, such as the semantic relations holding between the target word/expression and the set of foils within the proverb.

Since the ultimate goal of this work is to automatically generate the exercises, NLP tools and language resources are necessary to produce the exercises, in such a way that a large number of HIT be built, stored in a quiz database and made available to the student. We will ignore, in this paper, all technical aspects concerning the database construction and management, as well as the web interface and the procedures for correction, scoring, and keeping track of the students answers for future reference, and will focus solely on the linguistic aspects that building such games raises and the challenges they pose to a NLP, automatic generation approach. In Section 3, we presented some of the linguistic issues in choosing the appropriate distractors for a given target word, based on the semantic relations it may hold with other lexical items, using the information encoded in existing ontologies already built for Portuguese. We resume that exercise here, adding some further examples.

Consider the wellknown and already mentioned proverb, with the target word [*pássaro*]:

Mais vale um [pássaro] na mão que dois a voar

‘Better a bird in the hand than two flying’

Multiple, and complex, linguistic competences are at play and not all are necessarily under the scope of a HIT based on this sentence:

- the ability to **parse** the syntactic structure of a complex sentence with a subordinate comparative subclause (*mais C1 do que C2*);
- Parsing the main clause *C1* also involves: (a) the reconstruction of a reduced infinitive *Vinf w* = [*ter um pássaro na mão*] as the subject of *vale mais*; (b) the reconstruction of an indefinite (hence zeroed) generic subject of *ter* (*alguém/toda a gente ter um pássaro na mão* ‘for someone/anybody/everybody to have a bird in the hand’), this type of subject and its zeroing being responsible for the ‘generic’, that is, ‘proverbial’ value of the proverb;
- while the parsing of the subordinate *C2* involves: (a) the (lexical) anaphora resolution (*dois pássaros*), explained above; and (b) reductions of repeated material under the comparative conjunction *do que* ‘than’; an adequate analysis of the subordinate infinitive (*a Vinf = a voar*),
- as well as the knowledge of the **distributional constraints** on the subject of the verb *voar* in *C2* and of the complex clause underlying the relation between *pássaro* and *mão* in the first clause, involving locative preposition *em* and the particular grammatical value of *ter*.

It is on the latter, more semantically oriented, aspect that the type of HITs here suggested hinge, though all other factors be involved as well. The following distractors can be automatically selected using the semantic relations encoded in the ontologies and replacing the the target word by:

- a synonym (*ave*) or a quasisynonym (*piupiu*)
- a hypernym (^o*animal, vertebrado*)
- a diminutive of any of the words, including the target (*passarinho, avezinha*)
- a hiponym (*andorinha, arara, ...*)

All the examples above, except those marked with ^{oo}, can be directly retrieved from the databases.

The specific selection of the noun *pássaro* in the proverb and using its synonym *ave*¹⁰⁴ as a distractor may not be entirely adequate, as 5 instances of this proverb with *ave* were found with Google exact match search (restricted to the top domain **.pt**), against 120 (different) instances of the variant with *pássaro*. A check for the full string (with all the selected foils candidates, such as *ave*) in the web would prevent an intelligent system to produce incorrect foils (an

¹⁰³ https://en.wikipedia.org/wiki/Amazon_Mechanical_Turk#Artistic_and_educational_research

¹⁰⁴ We do not discuss here the issue of *pássaro* being a synonym, a hypernym or a hyponym of *ave*, which is a matter of linguistic adequacy of the semantic description, and pragmatically assume as correct any information in the language resources as given.

alternative correct answer). Notice the issue does not rise with *piupiu* and all other remaining options (no matches were found). Thus, such web search filtering mechanism should be of general use in any automatic exercise generation system. The choice of hypernyms conflicts with the somewhat constraint distribution found for the subject of *voar*. Besides, usually this strategy is not very productive, not only because of the limited number of lexical items functioning as hypernyms in the ontology hierarchy. The use of diminutives focuses not only on a morphologic competence, but also on the narrow distribution (or *frozenness*) of proverbs in general. It may be a good strategy for selecting foils, as diminutives (especially *inho* and *ito*) are a productive lexical device in Portuguese, which constitutes a further reason to be trained and learnt. Again, the filtering based on a web search would allow to discard *passarinho* (13 instances in Google) against *avezinha* (no exact match found). Finally, the list of hyponyms being very large would enable a substantial set of alternative foils.

Combining the different strategies in the same HIT could enable a system to produce a large number of HITs, all based on the same proverb. Irrespective of the computational complexity involved, all the necessary language resources and NLP tools exist and are adequate.

Another venue of research could result from exploring the fact that many proverbs are based on opposing/contrasting meaning, that is, words in an *antonymy* relation. However, in order to be able to explore such feature, one must first detect such proverbs having to do with antonym words.

To do so, we used the 388 word pairs of adjectival antonyms (e.g. the adjectives *direito/torto* ‘right, straight’/‘crooked, twisted’) encoded in PAPEL (Oliveira *et al.* 2008), and queried all the entries in the proverbs database containing such pairs, irrespective of their order: 556 instances of proverbs (and variants) were found in the database, which correspond to 330 different strings. Below are ten of the first proverbs (sorted alphabetically) of that list, involving the pairs of adjectives *fácil/difícil* ‘easy/difficult’, *certo/incerto* ‘certain’/‘uncertain’, *feliz/infeliz* ‘happy’/‘unhappy’, *amigo/inimigo* ‘friend’/‘enemy’ and *possível/impossível* ‘possible’/‘impossible’

- (1) *A crítica é fácil, a arte difícil.* ‘Criticism is easy, art (craft) is difficult’
- (2) *A crítica é fácil... a arte (é) difícil.* ‘Criticism is easy, art (craft) is difficult’
- (3) *A hora é incerta mas a morte é certa.* ‘The hour is uncertain but death is certain’
- (4) *A hora é incerta, mas a morte é certa.* ‘The hour is uncertain but death is certain’
- (5) *A inconstância da fortuna assusta os felizes e anima os infelizes.* ‘The fickleness of fortune scares the happy and animates the unhappy’
- (6) *A lisonja faz amigos e a verdade, inimigos.* ‘Flattery makes friends and truth enemies’
- (7) *A marido ausente, amigo presente.* ‘To absent husband, present friend (=lover)’
- (8) *A melhor maneira de nos desfazermos de um inimigo é fazer dele um amigo.* ‘The best way to discard an enemy is to make a friend out of him’
- (9) *A morte é certa, a hora incerta.* ‘Death is certain, the hour uncertain’
- (10) *A quem busca o impossível, justo é que até o possível se lhe negue.* ‘To whom that seeks the impossible, it is fair that even the possible be denied to him’

Notice, first, that some of these instances correspond just to graphic (punctuation) variants of the same paremiological unit, namely (1) and (2) or (3) and (4); secondly, the reversal of the order of the antonymic word pair yields certain variants (9) where the two parts of the proverbs have just been permuted (3)(4).

Let us consider again the proverb *Deus escreve direito por linhas tortas* ‘God writes straight with crooked lines’ and try to define possible HITs for this type of antonymy-based proverb. A very simple exercise would be to swap the adjectives:

Deus escreve torto por linhas direitas ‘God writes crookedly by straight lines’

The purpose of such foil could be, for example, to identify illformed proverbs among correct variants. However, proverbs involving antonymy are very prone to creative reuse, which explains why 12 instances of such form have been found in the web, including a very wellknown dramatic text¹⁰⁵:

Manuel: *Não é de espantar. Deus escreve torto por linhas direitas. Não é assim que se devia dizer?*

¹⁰⁵ http://storamjoao.blogspot.pt/2008/11/anlisedefelizmentehluar_29.html

‘Manuel: That does not surprise anyone. Gog writes crookedly by straight lines. It is not like this how one should say it?’ (Luís de Sttau Monteiro, *Felizmente há luar!*)

It is also noteworthy that, if this permutation were performed, many of the proverbs of the list above, they would still look like proverbs, though the interpretation of some of them would become awkward:

A crítica é difícil, a arte fácil. ‘Criticism is difficult, art (craft) is easy

A hora é certa mas a morte é incerta. ‘The hour is certain but death is uncertain’

A inconstância da fortuna assusta os infelizes e anima os felizes. ‘The fickleness of fortune scares the unhappy and animates the happy’

A lisonja faz inimigos e a verdade, amigos. ‘Flattery makes enemies and truth friends’

A marido presente, amigo ausente. ‘To present husband, absent friend (=lover)’

A melhor maneira de nos desfazermos de um amigo é fazer dele um inimigo. ‘The best way to discard an enemy is make a friend out of him’

A quem busca o possível, justo é que até o impossível se lhe negue. ‘To whom that seeks the possible, it is fair that even the impossible be denied to him’

On another perspective, one could also use the list of synonyms associated to each key adjective:

- **direito** (16 synonyms: *aprumado, desempenado, directo, erecto, especado, franco, imparcial, íntegro, justo, leal, liso, plano, rectiforme, recto, sincero, and vertical*)
- **certo** (19 synonyms: *ajustado, cabal, certo, claro, compassado, convencido, convicto, correcto, exacto, fixado, incontestante, indubitável, infalível, legal, positivo, preciso, seguro, tranquilo, and verdadeiro*)
- **torto** (16 synonyms: *desleal, embriagado, emviado, errado, esconso, esquelhado, estrábico, inclinado, injusto, oblíquo, torcido, torso, tortuoso, troncho, vasqueiro, vesgo*)
- **travesso** (32 synonyms: *amarotado, atratantado <sic>, atravessado, colateral, desenvolto, diabólico, diábril, endemoninhado, endiabrado, estouvado, gabiru, gaiato, garoto, inquieto, irrequieto, judio, levadinbo, levado, magano, maldoso, maroto, mexelbão, oblíquo, rabeador, rabino, remexido, roberto, safado, sapeca, traquinas, turbulento, vivo*)

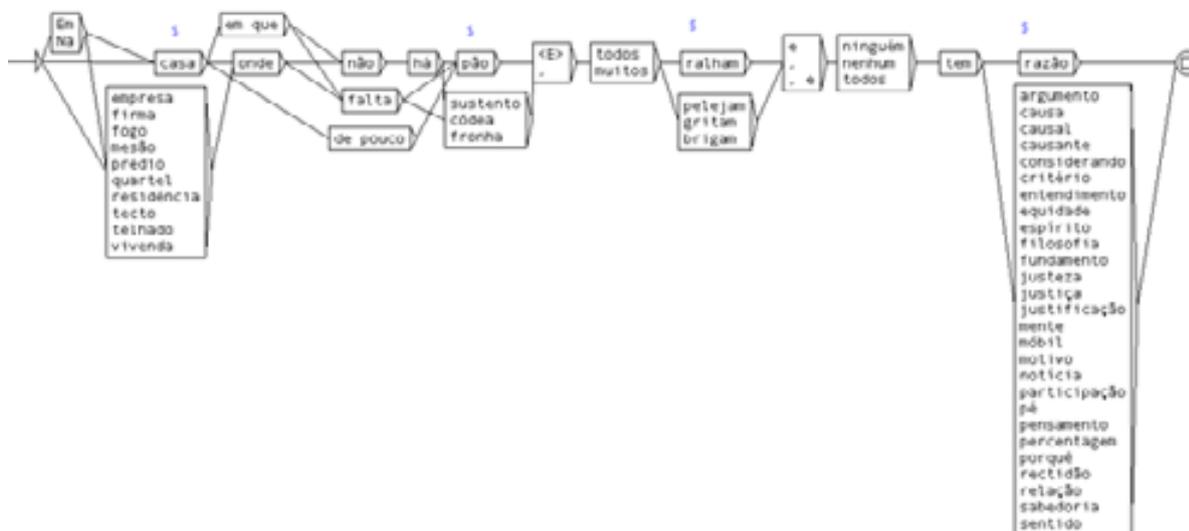
To create new distractors, it suffices to replace the key adjectives, one at time, keeping the others constant, by their synonyms. This relatively simple permutation would yield 3,260 distractors (though not all were correct), that would enable the student to practice his/her semantic competence about the antonymy relations each adjective may (or may not) hold with the other adjectives in the proverb.

This, however, may be different depending on the language variety one considers. For example, one can imagine that in Brazilian Portuguese (but not in standard European Portuguese!) the equivalence of the (manner) adverbial use of the adjectives *direito* = *reto* and eventually *direto* would lead to consider as natural the variants (we found instances of both in the web):

Deus escreve reto/direto por linhas tortas ‘God writes straight by crooked lines’

Another type of exercise could involve the simultaneous variation of more than one word at the same time, which would increase the level of difficulty of the HIT. For example, consider the proverb *Em casa onde não há pão, todos ralham e ninguém tem razão* ‘In a house without bread, everyone argues and no one is right’. Defining as target words the nouns *casa* ‘house’, *pão* ‘bread’ and *razão* ‘reason’ and the verb *ralhar* ‘argue/discuss/quarrel’, and using their equivalent words, marked as synonyms in PAPEL, would yield the combinations (several thousands) represented in the graph of Fig. 4.

Figure 4: Multiple word variation in a single proverb: synonyms associated to four target words in the proverb *Em casa onde não há [pão], todos [ralham] e ninguém tem [razão]* ‘In a house without bread, everyone argues and no one is right’.



Notice that for each noun, the semantic relations between the target word and the so-called equivalent word are not always the same, depending on the word sense. By exploring this sense variation, the student may improve his/her lexical competence, extending the vocabulary not just in the number of new words, but also in the number of senses he/she can associate to already known words. For example, for *casa* ‘house’, the professional sense of *empresa* and *firma* ‘company’ is not to be confounded to the sense of type of habitation building like *vivenda* ‘villa’ or *prédio* ‘apartment building’, nor with the figurative use of *tecto* and *telhado* ‘roof’, nor even with the almost technical value associated to *fogo* (literally: ‘fire’) ‘habitation’. Some of these “equivalent” words are just plainly bizarre, probably an error resulting from the automatic methods used to build the ontology and the lack of human verification (e.g. *pão* ‘bread’ and *fronha* ‘sleeping pillow cover’). Notice the figurative use of *pão* ‘bread’ in the equivalence with *sustento* ‘sustenance/livelihood’. The case of *razão* ‘reason’ is more interesting, since this use corresponds to a support verb (or light verb) construction, *ter razão* ‘to be right’, and the lack of equivalence results from the simple wordbyword equivalence (and the many senses the simple noun may present), without considering this multiword lexical unit. Probably, targeting this noun would result in uninterpretable HITs. However, very few NLP systems, nowadays, can parse support verb constructions adequately (Rassi *et al.* 2014b).

Looking beyond antonymy and synonymy, other semantic relations can be explored in the automatic generation of foils for a proverb, namely hyperonymy and hyponymy, or *wholepart* relations, i.e. holonymy and meronymy. An example of the later might be construed for the proverb:

Mais vale um pássaro na mão do que [dois (pássaros)] a voar

‘It is better to have a/one bird in the hand than two flying’,

where the numeral *dois* (pássaros) ‘two (birds)’ in the headless noun phrase could be replaced by *um bando* (*de pássaros*) ‘a flock (of birds)’ and then by different types of collective nouns, to explore the lexical knowledge about the distributional constraints on the use of collective nouns designating groups of animals (holonymy):

Mais vale um pássaro na mão do que um bando/cardume/cacho/chorrilho a voar

‘It is better to have a/one bird in the hand than a flock/school/bunch/ flying’

Naturally, we assume that the issue of anaphora resolution should have already been solved.

Finally another, rather challenging, exercise could involve discovering the appropriate proverb for a given context. However, instead of doing like some of the websites that explain the proverbs’ meanings by presenting examples produced by their authors, one would use NLP tools to retrieve from the web and other sources real, spontaneous instances of proverbs in their context of use. This requires being able to find proverbs in texts, especially the most frequent, and hence the better known, variants, and selecting those instances more prone to the generation of

adequate distractors. For example, one of the proverbs already mentioned above has been found in a blog¹⁰⁶, with an extensive enough context to provide some of the clues required for guessing it:

Estudos feitos revelam que mais de 50% dos conflitos conjugais acontecem devido a problemas financeiros. Há um ditado popular que diz “[Em casa onde não há pão todos ralham e ninguém tem razão]”, e será que não é verdade?

‘Studies show that more than 50% of marital conflicts happen because of financial problems. There is a popular saying that goes like “In the house where there is no bread all quarrel and no one is right,” and is it not true?’

In the preceding text, several nouns can serve as clues for this proverb: *conflitos conjugais* ‘marital conflicts’ and *problemas financeiros* ‘financial problems’. Other content words, like *estudos* ‘studies’, *verdade* ‘truth’ or even the verb *revelar* ‘reveal’, have little to do with the proverb’s lexical content. Furthermore, the proverb is clearly identifiable by the devices used to introduce it, namely, the quotation marks and the formula *Há um ditado popular que diz* ‘There is a popular saying that goes like’ (this type of formula, if identified, could be ruled out from the set of clues,

Different strategies can be put in place. A very simple approach is to retrieve from the *corpus* of proverbs the entries that have two or more of the same words of the contextual clues. In our *corpus*, there is a instance of such situation (though it is not clearly a proverb), where the word *problema* appears repeated (no case was found with any of the word pairs):

Os problemas existem para se lhes dar soluções; quem não quiser defrontar os trabalhos de soluções deve evitar os problemas.

‘Problems are there to be given solutions; who does not want to face the work required for the solutions should avoid any problems.’

Alternatively, from an ontology like MultilingualWordNet.PT¹⁰⁷ one can retrieve the words that have a given semantic relation with the contextual clue (Table 5), much as already shown above. As one can see, some words may be missing from the ontology. It is then relatively simple to extract from the *corpus* the proverbs containing two or more words from the combinations of these elements. For example, one finds:

A desgraça do homem é falar fino, esmorecer, brigar com a mulher e ficar perto.

‘The man’s misfortune is talking with a thin voice, fading/waning, fight with his wife and stay close.’

where the mention of a marital relation and the verb *brigar* ‘quarrel’ (but no mention of *problema* ‘problem’ nor *financeiro* ‘financial’) make this a good distractor. However, most of these word combinations may not exist in the *corpus*.

A more sophisticated approach would be to compute the word sense similarity (Clark *et al.*, 2010) between the words in the proverb, on one hand, and those of the context, on the other hand, thus retrieving from the *corpus* the proverbs that are similar/dissimilar with the values obtained.

Table 5: Semantic relations in MWN.PT

Word	[Semantic relation]
estudo	[hyperonym (is a kind of)]: <i>composição musical, música, peça musical</i>
conflito	[is a word for]: <i>briga, combate, luta</i> [hyperonym (is a kind of)]: <i>estado, acto colectivo (PE), ato coletivo (PB)</i> [hyponym]: <i>briga, choque, combate, contenda, desacordo, desavença, desentendimento, desinteligência, dissensão, divergência, duelo, embate, enfrentamento, fricção, guerra fria, lixe, luta, pancadaria, peleja, pugna, recontro, refrega, rixa</i>
conjugual	not found
financeiro	not found
problema	[is a word for]: <i>aperto, apuro, dificuldade, problema</i> [hyperonym (is a kind of)]: <i>acontecimento, sucedido</i> [hyponym]: <i>aflição, atrocidade, adversidade, desdita, desgraça, desventura, escândalo, indignidade, constrangimento, embaraço, inferno, interferência, massacre, ultraje</i>

¹⁰⁶ <http://reorganiza.pt/guiaorcamentofamiliarefcaz/>

¹⁰⁷ <http://mwnpt.di.fc.ul.pt/> (we disregard the synset structure).

7. CONCLUSIONS AND FUTURE WORK

In this paper, we present some available NLP tools and resources that can be used to develop didactic exercises, automatically generated and corrected, to train linguistic competence students of PFL (and also of native speakers) and built around proverbs. We demonstrated that existing exercises require an inordinate amount of time and effort to be sufficiently apt to train those competences, so the automation of the process of creating them, particularly in a gaming environment, would help students and teacher alike, and improve the learning process. We have shown that these existing exercises heavily rely on semantic relations among words. Portuguese already has a number of automatically built or manually crafted language resources, as well several NLP systems that could be use as the basis for such project. The iCALL framework provided by the REAP.PT project could be adopted for such project. Naturally, this would require a close interaction and collaborative work involving both linguists and computer scientist. Part of the preliminary work is already under way: a large *corpus* of proverbs and variants has been collected and the construction of tools for retrieving paremiologic units (proverbs and their variants) from text is being done. In the near future, we expect to present some results from this process.

ACKNOWLEDGMENTS

This work was partially supported by national funds through Fundação para a Ciência e a Tecnologia (FCT), ref. UID/CEC/50021/2013.

REFERENCES

- Bick, E. (2000). *The Parsing System "PALAVRAS": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Dr.phil. Thesis. Aarhus University. Aarhus, Denmark: Aarhus University Press.
- Bruckschen, M.; Muniz, F.; de Souza, J.G.C.; Fuchs, J. T.; Infante, K.; Muniz, M.; Gonçalves, P.N.; Vieira, R. and Aluísio, S. (2008). *Anotação linguística em XML do corpus PLN-BR*. Technical Report. São Carlos (SP).
- Chacoto, L.M.V.G. (1994). *Estudo e Formalização das Propriedades Léxico-Sintáticas das Expressões Fixas Proverbiais*. Master thesis, Faculdade de Letras da Universidade de Lisboa.
- Clark, A.; Fox, C. and Lappin, S. (eds.). (2010). *The Handbook of Computational Linguistics and Natural Language Processing*. Wiley-Blackwell.
- Correia, R.P.S. (2010). *Automatic Question Generation for REAP.PT Tutoring System*, Master Thesis, Universidade Técnica de Lisboa, Instituto Superior Técnico.
- Correia, R.; Baptista, J.; Mamede, N.; Trancoso, I. and Eskenazi, M. (2010). Automatic Generation of Cloze Question Distractors. In: *Proceedings of the Workshop on Second Language Studies: Acquisition, Learning, Education and Technology*, Tokyo, Japan.
- Correia, R., Baptista, J., Eskenazi, M. and Mamede, N. (2012). Automatic Generation of Cloze Question Stems. In *Proceedings of the 10th International Conference on Computational Processing of the Portuguese Language*. LNAI/LNCS 6001: 168-178. Berlin/Heidelberg: Springer.
- Costa, J. (1999). *O Livro dos Provérbios Portugueses*. Lisboa. Editorial Presença.
- Council of Europe. (2011). *Common European Framework of Reference for Languages: Learning, Teaching*. Council of Europe.
- Fellbaum, C. (1998, ed.) *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Dias-da-Silva, B.C. (2013). Modelagem linguístico-computacional de léxicos. In: LAPORTE, Éric; SMARSARO, Aucione; VALE, Oto A. (Orgs.). *Dialogar é preciso: linguística para processamento de línguas*. 1st edition, Vitória: PPGEL/UFES: 89-103.
- Gross, M. (1996). *Grammaire Transformationnelle du Français. Syntaxe de l'Adverbe*. Paris: ASSTRIL.
- González Rey, I. (2002). *La Phraséologie du Français*. Toulouse. Presses Universitaires du Mirail.
- Hoshino, A. and Nakagawa, H. (2005). A Real-Time Multiple-Choice Question Generation for Language Testing: A Preliminary Study. In *Proceedings of the Second Workshop on Building Educational Applications Using NLP, Association for Computational Linguistics*: 17-20.

- Machado, J.P. (1996). *O Grande Livro dos Provérbios*, 1st edition, Lisboa: Editorial Notícias.
- Mamede, N.; Baptista, J.; Diniz, C. and Cabarrão, V. (2012). STRING: An Hybrid Statistical and Rule-Based Natural Language Processing Chain for Portuguese. In: Caseli, H.; Villavicencio, A.; Teixeira, A. & Perdigão, F. (eds.), *Computational Processing of the Portuguese Language, Proceedings of the 10th International Conference, PROPOR 2012 Demo Sessions*, vol. Demo Session.
- Marques, J.S. (2013). *Anaphora Resolution*. Master thesis, Engenharia Informática e de Computadores, Instituto Superior Técnico, Universidade Técnica de Lisboa.
- Marrafa, P. (2001) *WordNet do Português: uma base de dados de conhecimento linguístico*. Lisboa: Instituto Camões.
- Maziero, E.G.; Pardo, T.; Di Felippo, A. and Dias-da-Silva, B.C. 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.
- Mel'čuk, I. and Polguère. (2007). *Lexique actif du français - L'apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français*. Bruxelles. De Boeck.
- Mitkov, R. (2002). *Anaphora Resolution*, Longman, Studies in Language and Linguistics.
- Molinier, C. and Levrier, F. (2000). *Grammaire des Adverbes: Description des Formes em 'ment'*. Droz. Genève
- Moreira, A. (1996). *Provérbios Portugueses*. Lisboa, Editorial Notícias.
- Oliveira, H. G.; Santos, D; Gomes, P. and Seco, N. PAPEL: a dictionary-based lexical ontology for Portuguese. In António Teixeira, Vera Lúcia Strube de Lima, Luís Caldas de Oliveira & Paulo Quaresma (eds.), *Computational Processing of the Portuguese Language, 8th International Conference, Proceedings (PROPOR 2008)* Vol. 5190, (Aveiro, Portugal, 8-10 de Setembro de 2008), Springer Verlag: 31-40.
- Oliveira, H.G. and 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.
- Palma, C. (2009). *Expressões Fixas Adverbiais Descrição léxico-sintáctica e subsídios para um estudo contrastivo Português-Espanhol*. Master thesis. Universidade do Algarve.
- Parente, S. (2005). *O Livro dos Provérbios*. 1st edition, Lisboa: Editora Âncora.
- Paumier, S. (2003). De la Reconnaissance de Formes Linguistiques a l'Analyse Syntaxique. Volume 2, Manuel d'Unitex. Ph.D. thesis, IGM, Université de Marne-la-Vallée.
- 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.
- Perkins, D. (1999). The many faces of constructivism. *Educational Leadership*, **5(3)**: 6-11.
- Pianta, E.; Bentivogli, L. and 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.
- Ranchhod, E.; Mota, C. and Baptista, J. (1999). A computational lexicon of portuguese for automatic text parsing. In *Proceedings of SIGLEX99: Standardizing Lexical Resources, 37th Annual Meeting of the ACL*: 74-80. College Park, Maryland, USA.
- Ranchhod, E.M. (2003). O Lugar das Expressões 'Fixas' na Gramática do Português in Ivo Castro & Inês Duarte (orgs.) *Razões e Emoção*, vol II. Lisboa: Colibri: 239-254.
- Rassi, A.; Baptista, J. and Vale, O. (2014). Automatic Detection of Proverbs and their Variants. Leibniz (Germany): Schloss Dagstuhl, Leibniz, Zentrum für Informatik, Dagstuhl Publishing: *Proceedings of the Symposium on Languages, Applications and Technologies (SLATE'14)*: 235-249.
- Rassi, A.P.; Santos-Turati, C.; Baptista, J.; Mamede, N. and Vale, O. (2014b) The fuzzy boundaries of operator verb and support verb constructions with dar "give" and ter "have" in Brazilian Portuguese. *Workshop on Lexical and Grammatical Resources for Language Processing (LG-LP 2014), COLING 2014*, At Dublin, Ireland: 92-101.

- Reis, S. (*in preparation*). *Constituição de um corpus de provérbios e variantes para representação da variação formal de unidades paremiológicas* (Technical Report). Faro: UALG.
- Rocha, P. and Santos, D. (2000). CETEMPúblico: Um *corpus* de grandes dimensões de linguagem jornalística portuguesa. In: Nunes, M.G. et al., eds., *V Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR'2000)*, São Paulo: ICMC/USP: 131-140.
- Silva, J.; Branco, A.; Castro, S. and Reis, R. (2010). Out-of-the-Box Robust Parsing of Portuguese. In *Proceedings of the 9th International Conference on the Computational Processing of Portuguese (PROPOR'10)*: 75-85.
- Simões, A. and 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.
- Vicente, A. (2013). *LexMan: um Segmentador e Analisador Morfológico com Transdutores*. Master thesis. Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, Portugal.
- 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.

Capítulo 3

O uso de provérbios no ensino de português

Artigo 3

Publicado em: Soares, R. & Lauhakangas, O. (Eds.) *10th Interdisciplinary Colloquium on Proverbs, Actas ICP16 Proceedings*. Tavira: AIP-IAP, 2017, pp. 521-538

Para uma melhor visualização das Figuras 3 e 4 deste artigo (p. 534 [p. 88]) colocámos aqui esta imagem.

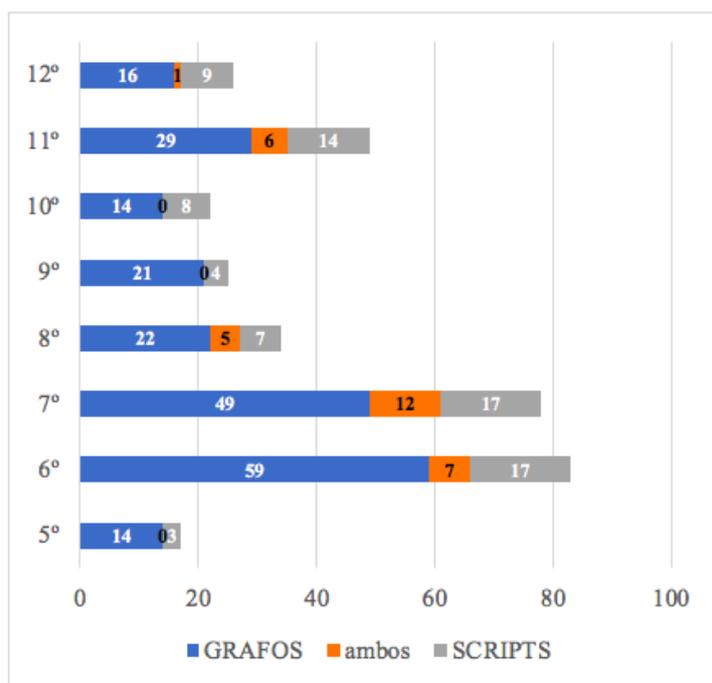


Figura 3, Provérbios encontrados por ano escolar

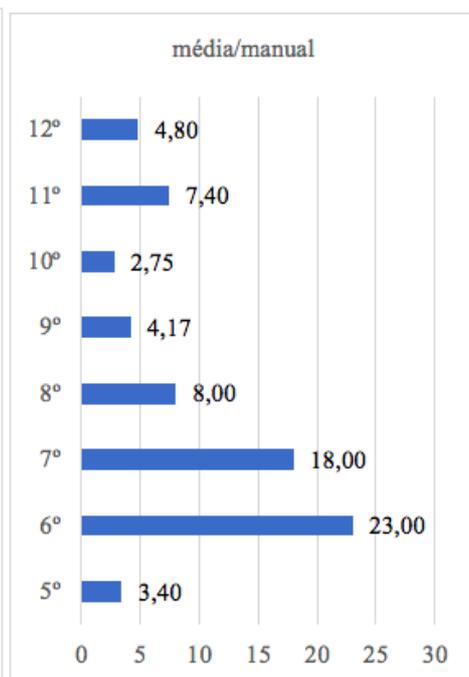


Figura 4, Média de provérbios por manual

Figura 3.1: Figuras 3 e 4 a cores

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 momentos do percurso escolar são os provérbios referidos tanto nos textos ordenadores como nos manuais escolares, bem como determinar os fatores 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 estudos ¹, 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.

Nos 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, dos 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 *Portfolio 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.

2.3. 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:

- para introduzir o tópico, o autor coloca a questão aberta: “Conhece algum provérbio em Português?”;
- 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”;
- 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”;
- a partir deste exercício, o aluno deverá selecionar três provérbios e apresentar outros três com sentido equivalente ou oposto;
- num outro exercício pede-se ao aluno que identifique os provérbios cujo conteúdo está representado sob a forma de uma imagem;

- é apresentado um exercício de palavras-cruzadas que tem alguns provérbios e cujo objetivo é completar a palavra em falta nesses provérbios;
- 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.
 - i. 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.

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.

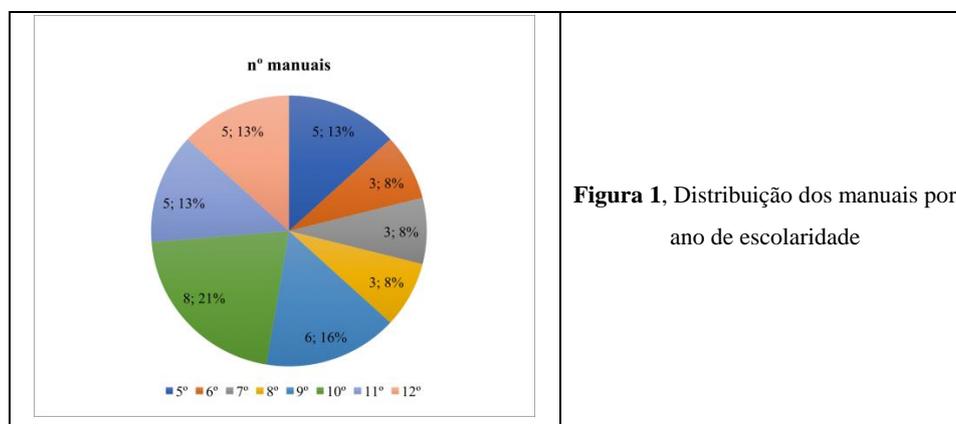


Figura 1. Distribuição dos manuais por ano de escolaridade

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: género-número, Verbo: pessoa-número-tempo-modo) e certas inserções.

Como resultado desta pesquisa obtivemos 239 ocorrências. Algumas destas são meras referências 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 primeiro
 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 apetecendo-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 toutiço 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*.

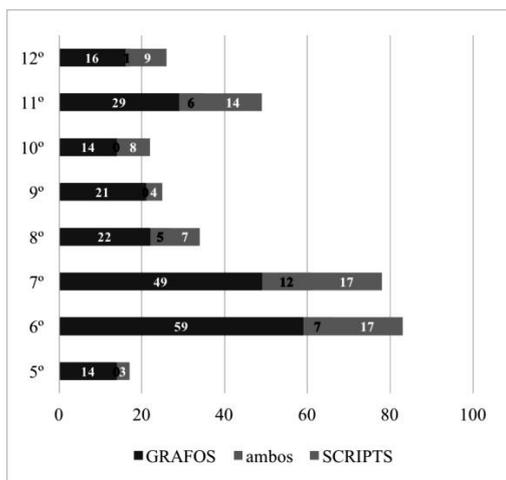


Figura 3, Provérbios encontrados por ano escolar

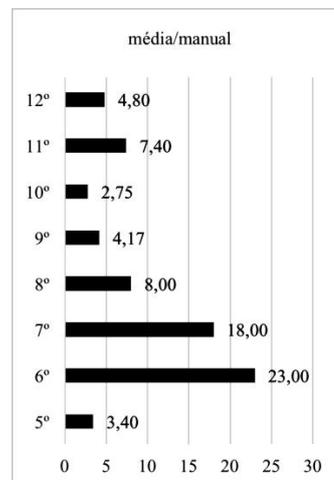


Figura 4, Média de provérbios por manual

Por outro lado, não obstante a distribuição do número de manuais por ano de escolaridade não ser homogênea 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 vezes ⁷, 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 UAIG 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 & GASPAS, Regina. (2011). *Programa de Português L2 para Alunos Surdos - Ensinos 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). Over-view 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.

Curricula Vitæ

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 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.

Emails: reis.soniamm@gmail.com, jbaptis@ualg.pt

----- |||| -----

THE USE OF PROVERBS IN THE TEACHING OF PORTUGUESE

Sónia REIS, Universidade do Algarve, *Portugal*

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

Abstract

Proverbs as linguistic elements marked by its widespread use and they are very rich structures, both from a cultural and a linguistic point of view, hence its potential as an important contribution to language teaching. This fully justifies the inclusion of proverbs in the official Syllabus and in the Curriculum Goals for Portuguese in our educational system. Although proverbs are only explicitly mentioned in the syllabus of the 5th and 6th grades of Portuguese basic education, they appear disseminated in textbooks until the 12th grade. In this work, we intend, at first, to verify at which school levels are proverbs explicitly referred to, both in the official syllabuses and connected documentation and in a sample of school textbooks, as well as to identify the factors that promote their use in such contexts. Secondly, we intend to determine which are the most widespread proverbs (and their variants) in that same sample of school textbooks, using natural language processing tools. The results will provide insight on the frequency and distribution of proverbs in textbooks and allow us to build a clearer picture of how proverbs are used in a language teaching context and the potential impact this representation may have in their respective lexical availability.

Key-words: lexical availability, language teaching, textbooks, natural language processing (NLP), Portuguese proverbs.

Email: reis.soniamm@gmail.com, jbaptis@ualg.pt

Capítulo 4

Os provérbios em manuais de ensino de português língua não materna

Artigo 4

Publicado em: Vlória Pinheiro & Gustavo Henrique Paetzold (Eds.) *Proceedings of Symposium in Information and Human Language Technology*. Uberlandia, MG, Brazil, October 2-5, 2017, Sociedade Brasileira de Computação, pp. 247-255

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, jbaptis@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 orientadores, no panorama europeu, o *Quadro Europeu Comum de Referência para as Línguas* (QERL) (Conselho da Europa, 2001) e o *Portifólio 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.

Os Provérbios em manuais de ensino de Português Língua Não Materna

é, um provérbio com as respetivas variantes) provérbios e variantes que ocorriam em mais do que uma coletânea. De seguida, representou-se sob a forma de transdutores de estados finitos (ing. *finite-state transducers*, FST) cada unidade paremiológica, começando pelas chaves que apareciam mais vezes repetidas, e prosseguindo por ordem decrescente de número de ocorrências da chave. Quando aplicados a um texto, estes transdutores permitem a identificação com um alto grau de precisão dos provérbios e respetivas variantes, admitindo variação morfossintática dos seus elementos, variação quanto aos sinais de pontuação e até um conjunto controlado de permutas, inserções e truncagens. Até ao momento, foram representadas sob a forma de grafo 133 unidades paremiológicas diferentes, que ao serem aplicados à base de dados de provérbios nos permitiram recuperar 2196 entradas (provérbios).

Dada a extensão da base de dados, este processo (manual) de formalização ainda está em curso, pelo que foi complementado, no imediato, com outros métodos, descritos em §3.2 e §3.3.

3.1. Pesquisa com transdutores representando unidades paremiológicas

Este método consistiu, como se disse atrás, na construção de FST, representando cada unidade paremiológica e associando as suas respetivas variantes. Na Figura 1 representa-se a variação formal encontrada no provérbio *Cão que ladra não morde*.

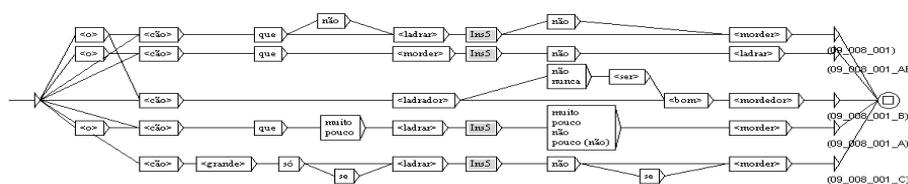


Figura 1. Transdutor de estados finitos representando a variação formal do provérbio *Cão que ladra não morde*.

Este grafo permite agrupar um grande número de variantes a uma mesma unidade paremiológica e captura os provérbios reconhecidos em toda a sua extensão nos textos em que se encontram. Um código alfanumérico convencional no nó de saída do grafo (à direita) identifica de forma unívoca o provérbio e o tipo de variante descrito. A aplicação deste tipo de transdutores ao *corpus* permitiu identificar corretamente 25 ocorrências (Figura 2), que representam 14 unidades paremiológicas diferentes, sem quaisquer resultados espúrios, isto é, todas as ocorrências encontradas eram efetivamente provérbios.

4. Bem diz o povo: «[Abril, águas mil\(07 002 001\)](#)». 5. O Steve matou saudades da língua dele. 168 Esc
 su. Bem diz o povo: «[Abril, águas mil\(07 002 001\)](#)». No entanto nem tudo foram desgraças. O tempo melh
 não vê corações. b) [As aparências iludem\(03 566 001\)](#), d) Quem avisa, amigo é. 132 (cento e trinta e
 stores públicos e... [casa onde não há pão, todos ralham e ninguém tem razão\(06 103 001\)](#). \ Tô: Por ou
 scutir os assuntos: "[da discussão nasce a luz\(03 3223 001\)](#)". Com os teus colegas, escolhe um tema que
 ido deste provérbio "[da discussão nasce a luz\(03 3223 001\)](#)". Conheces algum outro provérbio português
 i ouviram dizer que "[gordura é formosura\(05 488 001 A\)](#)»? Bom e estes jantares em que nos reunimos men

Figura 2. Excerto da concordância resultante da pesquisa com FST representando unidades paremiológicas

Nas secções seguintes, descreveremos técnicas complementares, usando outro tipo de transdutores, para a identificação automática de provérbios em textos.

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

Os Provérbios em manuais de ensino de Português Língua Não Materna

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

Os Provérbios em manuais de ensino de Português Língua Não Materna

provérbios encontrado 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 de 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 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.

Os Provérbios em manuais de ensino de Português Língua Não Materna

- 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*. 1ª edição, Lisboa, Ed. Notícias.
- Moreira, A. (1996). *Provérbios Portugueses*. Lisboa, Ed. Notícias.
- Parente, S. (2005). *O Livro dos Provérbios*. 1ª 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), 2ª edição. Lidel, Lisboa.

Os Provérbios em manuais de ensino de Português Língua Não Materna

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).

Capítulo 5

Estimating Lexical Availability of European Portuguese Proverbs

Artigo 5

Publicado em: Mitkov R. (Ed.) *Computational and Corpus-Based Phraseology. EUROPHRAS 2017. Lecture Notes in Computer Science*, vol 10596. Springer, Cham, pp. 232-244

DOI: https://doi.org/10.1007/978-3-319-69805-2_17

Publisher Name: **Springer, Cham Print ISBN 978-3-319-69804-5**

Online ISBN: **978-3-319-69805-2**

eBook Packages: **Computer Science**

Estimating Lexical Availability of European Portuguese Proverbs

Sónia Reis¹ and Jorge Baptista^{1,2}

¹ University of Algarve, Campus de Gambelas, 8005-139 Faro, Portugal
reis.soniamm@gmail.com, jrbaptis@ualg.pt

² L2F/INESC-ID, Lisbon, Portugal

Abstract. This paper relates data on lexical availability with data on textual frequency of proverbs in European Portuguese. Each data source should provide different perspectives on the use of proverbs in the language. This should allow an empirically well-motivated selection of proverbs aiming at the development of NLP resources, specifically for applications for learning Portuguese as a Foreign Language and for the diagnosis/therapy of speech impairments/disabilities. A large database (over 114,000 proverbs and their variants) was independently classified by two annotators, according to intuitively estimated lexical availability. Next, a random, stratified sample was selected and lexical availability was then confirmed with an online survey. Frequency data was gathered from two web browsers and a large-sized, publicly available, *corpus* of journalistic texts. Results from the survey, the web and the *corpus* by and large confirm the initial intuitive classification and a core of commonly used proverbs was defined.

Keywords: European portuguese proverbs · Frequency in *corpus* · Lexical availability

1 Introduction: Using Proverbs

Proverbs are used frequently and in many communicative contexts [1]. In spite of their colloquial/popular status and (mostly) oral transmission process [2], they are found not only in oral communication but also, though perhaps less frequently, in written discourses, and in different text types and genres [3], serving several types of rhetorical functions within discourse [4]. Because of their rich cultural and linguistic content [5], proverbs have been used in many applications, namely, as linguistic material for language learning [6] and language impairment diagnosis or speech therapy [7].

For language learning, proverbs provide a wide spectrum of expressive effects and a cross-cultural perspective on the language community and their symbolic heritage [2]. Furthermore, they are concise, often highly figurative, linguistic structures, naturally yielding to syntactic as well as culturally-oriented exploration in pedagogic context [8, 9].

As linguistic material for diagnosis/therapy of language impairment, particularly in the case of pathologies resulting from trauma, proverbs are deemed as effective tools, since they can be used as stimuli or prompts to exercise different cognitive structures, particularly long-term memory, even when the ability to speak is impaired [10–14].

Therefore, proverbs have been used in speech therapy, for example in tasks requiring the patient to complete a proverb, to explain its meaning, or the conditions in which it would be adequate to use it.

In spite of the crucial role that the adequate use of proverbs plays in the development of such exercises or tasks, little is said about their selection. Particularly in the case of didactic games involving proverbs, many available exercises could perfectly have used other linguistic material and no proper justification of the selected proverbs is provided [15].

Concerning proverbs, the selection of adequate material by specialist from these two areas faces several difficulties [16]. Selected proverbs should be commonly known, in order to reflect in a representative way, the culture of the language community using them. On the other hand, vocabulary involved in those expressions cannot be too rare or unknown. Furthermore, the choice of phraseological material should provide relevant items for the pedagogic goals of the exercises/activities [15].

As stimuli for speech therapy or language impairment diagnosis, selection of adequate examples is crucial. It is not easy to determine whether the fact of the patient not recognising the proverb, or the failure in producing an adequate answer to a fill-in-the-blank task, is due to a pathologic condition (e.g. dementia), or it is just due to the fact that he/she does not recognise/know the proverb.

Concerning Portuguese, and more specifically European Portuguese, [17] report several on-line resources, already available for learning Portuguese as a Foreign Language (PFL). The *Ciberescola da Língua Portuguesa*¹ and the *Centro Virtual Camões*² make available a set of didactic games, some of them involving proverbs. These games consist, basically, in completing a proverb or explaining its meaning.

For speech therapy/diagnosis, on the other hand, there are very few, publicly available, virtual therapists. One of them is the VITHEA system (Virtual Therapist for Aphasia Treatment) [18], which aims at the treatment of aphasia, featuring different types of visual and auditory stimuli, especially for eliciting vocabulary. To date, only a small number of exercises with proverbs have been produced.

Lexical availability is the key concept concerning these selection requirements. Besides other issues that may be task- or domain-specific (that should be used either in language learning or in speech therapy), a lexically available set of proverbs, consists of expressions:

- (a) that are easily recognised as such by the linguistic community as whole;
- (b) whose meaning and pragmatic conditions of use are widely known by native speakers;
- (c) that occur in a broad set of communicative situations.

Linguistic items presenting such requirements are considered to be lexically available, in the sense that they are part of the shared knowledge of the linguistic community, in as much the same way as the meaning and syntax of a commonly used verb is known, with high likelihood, by any native speaker of that given language. Lexically available

¹ <http://www.ciberescola.com/>, last accessed 2017/05/13.

² <http://cvc.institutocamoes.pt/>, last accessed 2017/05/13.

items are also deemed to show a significant frequency on the language daily use, so that frequency can be viewed as an indirect signal of that availability.

The problem at hand is, thus, a question of devising the appropriate method for selecting a representative sample from the *corpus* of proverbs available for a given language. This problem is somewhat similar to the definition of a common vocabulary from the large lexicon of a language [19]. However, because of their primary oral mode of transmission and their colloquial nature, finding evidence of proverbs' use in written *corpora* is not a trivial task: in several languages, such as Portuguese, many large-sized available *corpora* are built from journalistic texts, and writing and style conventions strongly advise against using such colloquial expressions [20]³. Besides, proverbs often present lexical and syntactical *variation* [21], which renders the task of finding them in texts much more complex than just looking for simple lexical items (words or phrases).

In view of the above, this paper aims at establishing, on solid empirical grounds, a core set of very commonly used proverbs, widely recognised (and adequately interpreted) by the majority of the (European) Portuguese, native speaking community. This selection could then be used in different scenarios, such as in language teaching and speech therapy, among other practical applications.

The remainder of the paper is structured as follows: Sect. 2 presents the method for manually selecting an initial set of proverbial expressions; then, using a stratified sample, classified according to their estimated lexical availability, that selection was validated using a survey. Next, in Sect. 3 the frequency of that sample is obtained from two popular web browsers, in order to further validate the initial selection. Finally, in Sect. 4 those same proverbs were queried in a large-sized, publicly available, European Portuguese *corpus*. Frequency data from the two types of sources (web and *corpus*) are compared against the estimated lexical availability, in order to produce an empirically well-motivated selection of commonly used proverbs that may be reliably used for different applications. The paper concludes (Sect. 5) by presenting the main findings and suggestions for future work.

2 Lexical Availability and Proverb Selection

In order to tackle the selection of an initial set of proverbs, candidate to the status of lexically available items, a large data base with 114,413 proverbs and their variants was used [22]. These were collected from four dictionaries of European Portuguese, which were digitised and then manually corrected. Each proverb was given a unique identifier (ID), indicating its source. After removing the stop words, each proverb was associated to a set of keywords: full verbs, nouns and adjectives, for the most part.

Two annotators both native speakers of European Portuguese and extensive knowledge of the proverbial stock of the language independently marked the proverbs they recognised and deemed as usual. Annotator 1 marked 739 proverbs, while Annotator 2 marked 379 proverbs. This produced a tiered list of proverbs, ranked by levels (0 to 2).

³ http://static.publico.pt/nos/livro_estilo/13-rigor-e.html, last accessed 2017/05/13.

In total, 276 expressions were considered usual by both annotators (level 2), 566 proverbs were considered usual by only one of the annotators (level 1), and the remainder forms (113,571) have not been marked by either (level 0).

The initial assumption is that level 2 proverbs are lexically highly available expressions, level 1 are less so (only moderately available), and level 0 items, constituting the bulk of the database, though they are part of the *corpus* of proverbs of the language, are not sufficiently usual (seldom available) to be included in a selection aiming at the applications envisaged in this paper (language learning and speech therapy, for example).

Since only two annotators were involved, it was then deemed necessary to further confirm their selection. Furthermore, since a significant mismatch was found between the two annotations, it should be ascertained which proverbs from level 1 (where the two annotators did not agree) should be integrated in level 2, or left as only moderately available (level 1), or even removed from the selection altogether (and integrated in level 0).

A survey was thus built to confirm this initial selection. However, because the total set of proverbs (and variants) from levels 1 and 2 is too large to be presented in a survey to a wide audience, for it would require a long time to be answered by each participant; a random, stratified sample of the list of proverbs was produced, namely 50 items from levels 1 and 2 (25 from each), and 50 items from level 0. Because of the random selection, the items were manually revised in order to avoid repetition of proverbs from two different sources or using two variants of the same proverb.

Google Forms was used to build the survey and collect the answers. Some personal data was collected to characterise the sample: gender (M/F/undisclosed), age (less than 18, 18-30, 31-50, over 50), school level (basic, secondary, university level, other), nationality (short answer) county of residence (Portugal's 22 counties, including the autonomous regions of Madeira and Azores), area of residence (urban/rural). The detailed analysis of this data is presented elsewhere [16]. The selected proverbs were presented in random order, but always in the same order to every participant. For each proverb, the participant was asked to indicate whether he/she did not know proverb, or knew it but did not use it, or knew it and used it. These answers correspond to the 0 to 2 levels of the tiered selection of the random sample.

The survey was divulged among the authors' list of contacts, both individuals and groups, potentially reaching over 3,600 people. The survey was open for 7 days, and 735 answers were gathered before closing the survey (answer rate: 20,4%).

Figure 1 shows the results from the survey. It can be seen that most answers for the level 2 proverbs (ID-001 to ID-025) were recognised and marked by the subjects, either as in the same level 2 (known and used) or in level 1 (known but not used). A small number of cases do not follow this pattern: In the case of proverb:

ID-002 *Não se apanham trutas a bragas enxutas*
 'Trouts cannot be cached/fished with dry trousers'

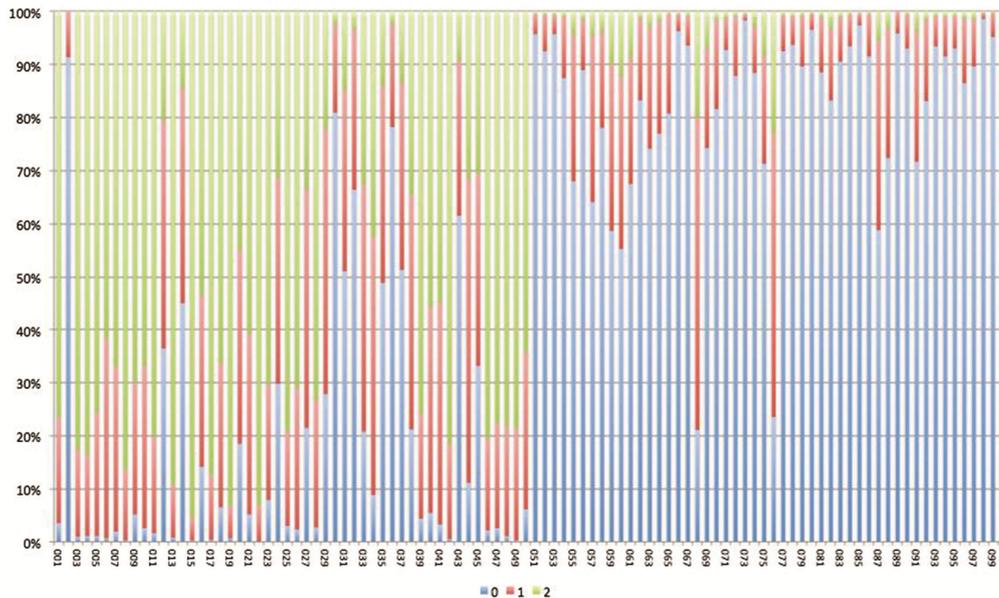


Fig. 1. Assigning lexical availability to proverbs. Horizontal axis: the sample of 100 proverbs, identified by their ID code; from ID-001 to ID-025: reference level 2 (highly available); from ID-026 to ID-050: reference level 1 (moderately available); from ID-051 to ID-100: reference level 0 (seldom available). Vertical axis: Percentage of answers from the survey ($N = 735$) as they map onto the reference values: 2: I know and use this proverb; 1: I know but I do not use this proverb; 0: I do not know this proverb.

most participants did not know the proverb and only 8% knew it but did not use it. This proverb includes the archaic word *bragas* ‘trousers’, which may be one of the reasons for it not being recognized. In fact, common variants of the proverb replace this disused noun either for the modern equivalent *calças* ‘trousers’ or by a phonetically similar noun *barbas* ‘beards’, which does not change the overall figurative meaning of the expression. For another four cases, the sum of answers indicating levels 1 and 2 is above 50%:

ID-012 *São mais as vozes que as nozes*
‘There are more voices than nuts’,

ID-014 *A preguiça morreu de sede à beira da água*
‘The sloth died of thirst (sitting) by/next to the water’,

ID-020 *A morte não escolhe idades*
‘Death chooses no ages’, and

ID-024 *Hoje por mim, amanhã por ti*
‘Today for me, tomorrow for you’.

In some cases, this can also be a result of the random selection of these proverbs/variants. In the case of proverb ID-020, a much more common variant exists with the noun *amor* ‘love’, which was confirmed by the queries on the web: the frequency ratio *mortelamor* ‘death/love’ in Google is 40/70 (0.57), and in Bing 40/237 (0.17).

On the other hand, among the many variants of proverb ID-024, there is one where 1st- and 2nd-person pronouns switch places:

Hoje por ti, amanhã por mim
‘Today for you, tomorrow for me’.

This variant (an instance of the so-called ‘golden rule’) is also almost as frequent as the one shown in the survey. As the survey’s variant constitutes an ‘inversion’ of the golden rule, this may be the cause for the lower availability level assigned by the survey. To sum up, selecting the most lexically available variants of a proverb proves to be almost as much important as choosing the proverbs themselves, in view of defining their lexical availability.

The situation is somewhat fuzzier in the case of level-1 proverbs (ID-026 to ID-050). Notice that this group of proverbs corresponds to a disagreement between the two annotators, as only one selected them as lexically available. Again, the (random) choice of a variant can be the cause for the proverb not being recognised. In the case of proverb:

ID-043 *Morra o gato, morra farto*
‘May the cat die, may it die fully satisfied’

the variant

Morra Marta, morra farta,

with the proper noun *Marta* ‘Martha’, is much more frequent. Queries on the web yielded a *gato/Marta* ratio of 3/38 (0.08) in Google and 2/59 (0.03) in Bing. The remaining cases caution for a careful review of the full list of level-1 proverbs (and their variants), using frequency data (from the web and eventually other sources) to support this classification.

Finally, all level-0 proverbs (ID-051 to ID-100) were also assigned the same lexical availability level by the subjects, except for two proverbs (ID-068 and ID-076), since a level 1 was assigned instead.

This data seems to confirm in general the manual assignment of 0- and 2-level of lexical availability to the sample of proverbs. The cases from level 1 must be considered with care, as results do not show a clear-cut distinction between this level and the other two.

3 Proverbs in the Web

The frequency of the same sample of proverbs was obtained from querying two web browsers, Google and Bing. The query was restricted to the exact matches, within the Portugal top domain (.pt), and selecting only pages written in European Portuguese. The matches were manually perused for false positives. Figure 2 shows these results. It is clear that the two web sources produced very similar results (Pearson correlation coefficient: 0.96), although the correlation is much higher for level-2 (0.94) and level-1 (0.91) proverbs, than for level-0 (only 0.86).

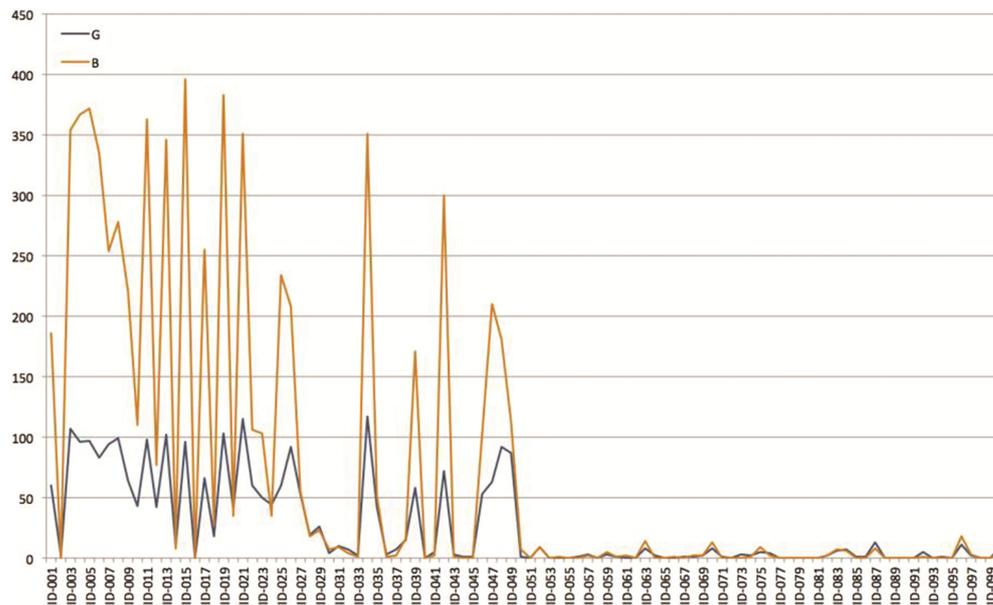


Fig. 2. Proverbs Frequency in two web browsers. Horizontal axis: the sample of 100 proverbs, identified by their ID code; from ID-001 to ID-025: reference level 2 (highly available); from ID-026 to ID-050: reference level 1 (moderately available); from ID-051 to ID-100: reference level 0 (seldom available). Vertical axis: hit counts, exact match, top domain Portugal (.pt), language: Portuguese (Portugal); values: Google (G, retrieved on 2017/05/03) and Bing (B, retrieved on 2017/05/13).

Table 1 shows the total and the average frequency of proverbs matched by the browsers, and the breakdown by reference level of lexical availability. It is noteworthy that Bing produced 2.7 times more matches than Google. When compared against the reference, the frequency values were slightly higher for the results from Google (Pearson: 0.73) than for those from Bing (0.69), while the sum of the frequency values from both browsers yield an intermediate value (0.70). This corresponds to a relatively high correlation.

Table 1. Total and average frequency of proverbs matched by the web browsers Google and Bing, and their sum (G + B), considering the entire sample (100 proverbs; 'All'), and by reference level ('L-2' to 'L-0').

Level	Total			Average		
	Google	Bing	G + B	Google	Bing	G+B
All	2,594	7,143	9,143	26	71	97
L-2	1,650	5,196	6,846	66	208	274
L-1	833	1,834	2,667	33	73	107
L-0	111	113	224	2	2	4

4 Proverbs in *Corpus*

Finally, the same proverbs were searched in the *CetemPúblico corpus* [23]⁴. This is a large-sized, publicly available *corpus* of journalistic text, collected from the online edition of the European Portuguese newspaper *Público*, and containing about 9,6 million words.

To process and query the *corpus*, the UNITEX linguistic development platform [24]⁵ was used. The *corpus* was processed using the European Portuguese language resources distributed with the system. The queries were carried out using finite-state transducers (FST) that are built using this platform formalism. These FST define a linguistic pattern to be matched and output the ID of the proverb corresponding to the matched string. Figures 3 and 4 show the FST used for querying the proverb *Santos de casa não fazem milagres* ‘Home saints don’t make miracles’.

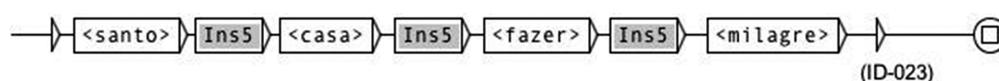


Fig. 3. FST built for querying the proverb *Santos de casa não fazem milagres* ‘Home saints don’t make miracles’ in *corpora*: sequences of keywords. Words inside chevrons represent lemmas. Grey boxes *Ins5* are subgraphs for insertions of 0 up to 5 words.



Fig. 4. FST built for querying the proverb *Santos de casa não fazem milagres* ‘Home saints don’t make miracles’ in *corpora*: graph describing lexical variants. Words inside chevrons represent lemmas. Grey box *Ins2* is a subgraph for insertions of 0 up to 2 words.

Two methods of querying were used:

- a set of FST describing the sequence of keywords (mainly nouns, verbs and adjectives) that characterise the proverb, allowing for a window of 0 to 5 words between them, and a small set of punctuation marks⁶: <;, ([] /and. .. >. This type of FST was automatically built from the database. Keywords are represented by their lemmas.
- another set of FST, manually built to describe all the variants of a given proverb found in the database, or deemed as reasonably probable to occur, according to the vocabulary involved and the syntactic structure of the proverb. These FST are only available for the 50 proverbs from levels 2 and 1, since level 0 expressions were not considered sufficiently relevant to be represented in this way.

⁴ www.linguateca.pt/cetempublico.

⁵ <http://unitexgramlab.org/>.

⁶ These insertions are reported in the sub-graphs - the grey boxes in Figs. 3 and 4.

Table 2 shows the results from applying the two sets of graphs to the *CetemPúblico corpus*. As expected, only some few instances (34) of the sampled proverbs were found in the *corpus*, corresponding to 13 different proverbs, from which 30 proverbs (instances) were matched by both methods. The keywords' graphs matched 40 instances (hence 10 false-positives), while the variants' graphs are very precise (no false-positive matches).

Table 2. Sampled proverbs matched in the *CetemPúblico corpus*. Column 'Var' indicates the number of matches using the proverbs' variants FST, while column 'Key' corresponds to FST with the keywords. An approximate translation of the proverbs (or an equivalent expression) is provided.

ID	Level	Proverb 'equivalent/Translation'	Var	Key
ID-003	2	<i>Depois da tempestade vem a bonança</i> 'After the storm comes the calm'	2	2
ID-013	2	<i>Quem sabe sabe</i> 'Who knows, knows'	4	4
ID-015	2	<i>Mais vale tarde do que nunca</i> 'Better late than never'	1	1
ID-019	2	<i>O tempo voa</i> 'Time flies'	1	1
ID-021	2	<i>O seguro morreu de velho</i> 'The careful man died of old age'	5	5
ID-022	2	<i>A esperança é sempre a última a morrer</i> 'Hope is the last to die'	3	3
ID-023	2	<i>Santos de casa não fazem milagres</i> 'Home saints don't make miracles'	7	7
ID-026	2	<i>O sol quando nasce é para todos</i> 'The sun when it is born is for everyone'	2	0
ID-034	1	<i>Nem só de pão vive o homem</i> 'Man does not live by bread alone'	2	2
ID-038	1	<i>Um crime não justifica outro</i> 'One crime does not justify another'	1	0
ID-042	1	<i>Quem avisa, amigo é</i> 'He who warns is a friend'	1	1
ID-049	1	<i>O tempo é dinheiro</i> 'Time is money'	4	4
ID-050	1	<i>Não degenera quem sai aos seus</i> 'He who resembles his own [people] does not degenerate'	1	0
		Total	34	30

The (small) difference between the results from the variants' graphs and from the keywords' graphs (proverbs ID-026, ID-038 and ID-050) is due to the fact that the variants' graphs include lexical variants of the keywords, while the keywords' graphs only consider one of those lexical variants.

For example, for proverb:

ID-026 *O sol quando nasce é para todos*
 ‘The sun when it is born is for everyone’

the keywords’ graph only considers the keywords *sol-nascer-ser-todos* ‘sun-rise- be-everyone’, but the corresponding variants’ graph allows *brilhar* ‘shine’ instead of *ser* ‘be’.

For proverb:

ID-038 *Um crime não justifica outro*
 ‘One crime does not justify another’

the lexical variant *erro* ‘mistake’ is allowed but it was not a keyword.

Finally, for proverb:

ID-050 *Não degenera quem sai aos seus*
 ‘He who resembles his own people does not degenerate’

the lexical variant *puxar* ‘pull/take’ was represented in the variants’ graph, but not in the keywords.

Only level-2 and level-1 proverbs were matched, 8 and 5, respectively. Though the number of instances is low, the most frequently occurring, level-1 proverbs, namely,

ID-049 *Tempo é dinheiro* (4 matches)
 ‘Time is money’

and (the fragment of) the biblical quote

ID-034 *Nem só de pão vive o homem* (2 matches)
 ‘Man does not live by bread alone’ (Mt 4:4)

may lead us to reclassify them as level-2.

Notice that the FST account for morphosyntactic variation of the proverbs’ keywords. For example, the gender-number variation of *santo* ‘saint’, or number variation on *milagre* ‘miracle’, as well as subject-verb agreement, are all handled by the graphs, using the lexical resources available with the system:

... *Portanto, para a Qualidade Total, santo de casa é quem faz milagres...*
 ‘Thus, for the Overall Quality, home saint (masc.-sg.) is the one that does miracles (pl.)’

... *E santa de casa não faz milagre...*
 ‘Home saint (fem.-sg.) doesn’t do miracle’

... *contrariando a tese de que santo de casa não faz milagre...*
 ‘against the thesis that home saint (masc.-sg.) doesn’t do miracle (sg.)’

The proverb:

ID-050 *Não degenera quem sai aos seus*
 ‘He who resembles his own people does not degenerate’

is more often used with the subject in the canonic word order, v.g.

Quem sai aos seus não degenera.

This variant had been used in the survey because of the random selection method used for the sampling. The more common variant, showing the basic word order was also checked. However – and quite surprisingly, no match was found in the *corpus*.

Contrasting with the low frequencies observed in this *corpus*, all these proverbs are quite frequent in the web (see Sect. 3). In average, these proverbs occurred 78 times in (Google) and 233 times (in Bing) – standard deviation, approximately 37 and 146, respectively; and totalling 1,017 and 3,034 respectively.

In spite of the low frequency observed in the *corpus*, it seems possible to conclude that:

- (a) most probably due to the journalist nature of the *corpus*, the occurrence of proverbs is this type of text is scarce, as expected;
- (b) even so, the lexical availability level, manually assigned to the proverbs' sample, has been confirmed, since no level-0 expressions were found, and there are more instances of level-2 than of level-1 proverbs.

5 Conclusion and Future Work

This paper set out to establish the lexical availability of a large-sized database of with over 114,000 proverbs. A preliminary estimation, carried out by two annotators, was confirmed by and large using data obtained through a survey (735 participants) and through queries on two popular web browsers (Google and Bing). Results from queries over a large-sized *corpus* of journalistic text also confirmed the initial expectations that the use of this type of linguistic expressions is often limited by style conventions to oral/colloquial communicative contexts.

In view of the results, it is reasonable to extend the lexically available status to all the level-2 manually selected proverbs. These will constitute the main core of a lexicon of commonly used proverbs. Level-1 proverbs, in general, and certain difficult or interesting cases, both from level-2 and level-0, will have to be studied further. In some cases, only some variants of a given proverb should be assigned level-2 status, while the remainder variants may be attributed to level-1 (moderately available) or even level-0 (seldom available). In other (rarer) cases, level-0 proverbs (or variants) may have to be raised to level-1, too.

The 50 finite-state transducers already built will now be extended to the remainder of level-2 proverbs and, time allowing, to level-1, after careful revision of this list. Eventually, surveying the lexically availability of the remaining entries from level-1 will be useful.

Previous experiments using a Brazilian Portuguese database of approximately 3,500 proverbs (614 types or paremiological units) over a relatively large *corpus* of this language variety [25] showed that the approach of using just the proverb's keywords (surface forms) is a reasonably effective strategy for detecting proverbs in texts.

In this paper, however, the keywords in the FST were (for the most part) lemmatized, which broadened the search area of the queries in the *corpus*.

A similar experiment has already been carried out on a *corpus* of Portuguese textbooks [15], but due to technical shortcomings of the linguistic platform used, it was not

possible to build FST with lemmatized keywords for the entire database. As a consequence, only the keywords' surface forms were used, narrowing the queries' search space. Having narrowed down in this paper the set of lexically available (or, at least, moderately available) proverbs of European Portuguese, it should now be possible to produce a new, richer resource, equivalent to that of [25], improving the accuracy and recall of proverb identification in texts.

More importantly, with this paper, an empirically motivated list of lexically available proverb (and variants) has now been produced⁷, which can be used in a reliable way to develop many types of applications, for example, for diagnosis/therapy of some speech disorders or for didactic games for language learning.

Acknowledgements. This work was partially supported by national funds through Fundação para a Ciência e Tecnologia (FCT) with reference UID/CEC/50021/2013.

References

1. Charteris-Black, J.: Proverbs in communication. *J. Multiling. Multicult. Dev.* **16**(4), 259–268 (1995)
2. Mieder, W.: *Proverbs – A Handbook*. Greenwood Press, London (2004)
3. Rezaei, A.: Rhetorical function of proverbs based on literary genre. *Procedia Soc. Behav. Sci.* **47**, 1103–1108 (2012). Elsevier Ltd.
4. Meira, A.: *Casa de ferreiro, espeto de pau: uma análise das relações retóricas a partir do uso dos provérbios como estratégia argumentativa em textos da internet*. (Ph.D. thesis), Faculdade de Letras da UFMG, Belo Horizonte (2015)
5. Hrisztova-Gotthardt, H., Varga, M. (eds.): *Introduction to Paremiology: A Comprehensive Guide to Proverb Studies*. DeGruyter, Berlin (2015)
6. Council of Europe: *Common European Framework of Reference for Languages: Learning, Teaching*. Council of Europe (2001)
7. Chaika, E.: *Linguistics, Pragmatics and Psychotherapy – A Guide for Therapists*. Whurr Publishers, London and Philadelphia (2000)
8. Arif, M., Abdullah, I.: The impact of output communication on EFL learners' metaphor second language acquisition. *Soc. Sci. (Pakistan)* **11**(9), 1940–1947 (2016)
9. Salbego, N., Osborne, D.: Schema activation through pre-reading activities: teaching proverbs in L2. *BELT Braz. Engl. Lang. Teach. J.* **7**(2), 175–188 (2016)
10. Gorham, D.: A proverb test for clinical and experimental use. *Psychol. Rep.* **1**, 1–12 (1956)
11. Benton, A.: Differential behavioral effects in frontal lobe disease. *Neuropsychologia* **6**(1), 53–60 (1968)
12. Gibbs, R., Beitel, D.: What proverb understanding reveals about how people think. *Psychol. Bull.* **118**(1), 133–154 (1995)
13. Siqueira, M., Marques, D., Gibbs Jr., R.: Metaphor-related figurative language comprehension in clinical populations: a critical review. *Scripta* **20**(40), 36–60 (2016)

⁷ This list is available to the scientific community at: https://www.researchgate.net/publication/319465467_List_of_100_proverbs_annotated_with_lexical_availability (DOI: 10.13140/RG.2.2.30110.64326).

14. Vas, A., Spence, J., Eschler, B., Chapman, S.: Sensitivity and specificity of abstraction using gist reasoning measure in adults with traumatic brain injury. *J. Appl. Biobehav. Res.* **21**(4), 216–224 (2016)
15. Reis, S., Baptista, J.: O uso de provérbios no ensino do Português (The use of proverbs in the teaching of Portuguese). In: Proceedings of the 10th Interdisciplinary Colloquium on Proverbs, Tavira, Portugal, November 6–13, 2016 (in print)
16. Reis, S., Baptista, J.: O provérbio como estímulo num terapeuta virtual (Proverbs as a stymulus of a virtual therapist). In: 6th 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, accepted for publication)
17. Reis, S., Baptista, J.: 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 (2016b)
18. Abad, A., Pompili, A., Costa, A., Trancoso, I., Fonseca, J., Leal, G., Farrajota, L., Martins, I.: Automatic word naming recognition for an on-line aphasia treatment system. *Comput. Speech Lang. Elsevier* **27**(6), 1235–1248 (2013)
19. Coxhead, A., Nation, P., Sim, D.: Measuring the vocabulary size of native speakers of English in New Zealand Secondary schools. *N. Z. J. Educ. Stud.* **50**, 121–135 (2015)
20. Martins, E.: Manual de redação e estilo, 3rd edn. O Estado de S. Paulo, São Paulo (1997)
21. Chacoto, L.: Estudo e Formalização das Propriedades Léxico-Sintáticas das Expressões Fixas Proverbiais, (Master thesis), Faculdade de Letras da Universidade de Lisboa, Lisboa (1994)
22. Reis, S., Baptista, J.: Portuguese proverbs: types and variants. In: Corpas Pastor, G. (ed.) *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives*, 208–217. Editions Tradulex, Geneva (2016)
23. Santos, D., Rocha, P.: Evaluating CETEMPúblico, a free resource for Portuguese. In: Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (Toulouse, 9–11 de julho de 2001), pp. 442–449 (2001)
24. Paumier, S.: *Unitex 3.1 User Manual*. Université de Paris-Est/Marne-la-Vallée – Institut Gaspard Monge, Noisy-Champs (2016)
25. Rassi, A., Baptista, J., Vale, O.: Automatic Detection of Proverbs and their Variants. In: Pereira, M., Leal, J., Simões, A. (eds.) *Proceedings of the Symposium on Languages, Applications and Technologies (SLATE 2014)*, pp. 235–249. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Leibniz (2014)

Capítulo 6

O provérbio como estímulo num terapeuta virtual

Artigo 6

Aceite para publicação em: *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

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 perspectivas 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 reportório 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 Rastreio 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

¹ <https://vithea.l2f.inesc-id.pt/>

frequente e intensivo às sessões de reabilitação. O VITHEA é uma plataforma on-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 dispuséssemos 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

² <https://goo.gl/forms/PH8uMlnhsFMIz2SQ2>

³ <https://www.google.com/forms>.

nessa amostra recorrendo a um número elevado de respondentes. Dado o tempo que a aplicação deste tipo de instrumentos implica 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 tanto 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 elicitación 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 consideração a estrutura formal, bem como as propriedades sintáticas, semânticas, lexicais, fonéticas dos provérbios, foram desenvolvidos *distratores* que apresentam diferentes relações com as palavras-

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.

chave dos mesmos. Essas relações são essencialmente de natureza semântica (hiperoníma/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 provérbios 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

⁵ www.censos2011.ine.pt

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] *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 seleccionadas 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*, 35, 554-559.
- Bhagal, 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.

Capítulo 7

Estudo das estruturas proverbiais com sujeito indefinido: Provérbio ou expressão fixa?

Artigo 7

Aceite para publicação em: Soares, R. & Lauhakangas, O. (Eds.) *12th Interdisciplinary Colloquium on Proverbs, Actas ICP18 Proceedings*. Tavira: AIP-IAP, 2018



ESTUDO DAS ESTRUTURAS PROVERBIAIS COM SUJEITO INDEFINIDO – PROVÉRPIO OU EXPRESSÃO FIXA?

Sónia REIS¹ & Jorge BAPTISTA^{1,2}, ¹Universidade do Algarve, ²INESC-ID, *Portugal*

Resumo

Os provérbios, tal como outras expressões fixas do português europeu (Baptista *et al.* 2016), apresentam diferentes estruturas sintáticas admitindo, por vezes, diferentes tipos de variação (García-Page, 2008; Reis, 2014; Steyer, 2015). A fim de melhor determinar a fronteira entre expressões proverbiais e expressões fixas, pretendemos aqui analisar as frases fixas com sujeito indefinido, que são usadas proverbialmente, como *Perca-se tudo menos a honra. / O Pedro perdeu tudo menos a honra.* e que permitem a sua integração no discurso. O objetivo é determinar os mecanismos sintáticos que permitem este uso proverbial de expressões fixas e assim estabelecer critérios mais precisos para a seleção do tipo de estruturas a integrar nas coletâneas de provérbios.

Palavras-chave: *Corpora*, classificação sintática, expressões fixas, Processamento da Linguagem Natural, provérbios.

Os provérbios, tal como outras expressões fixas (EFs) (Baptista *et al.* 2016), apresentam variadas estruturas sintáticas admitindo, por vezes, diferentes tipos de variação (García-Page Sánchez, 2008; Steyer, 2015). Associados muitas vezes a fórmulas introdutórias no discurso (*lá diz o ditado, como se costuma dizer...*) e compreendidos como enunciados completos ou textos autónomos dentro de um texto maior, com um valor de citação; os provérbios são ainda reconhecidos como tal por apresentarem, geralmente, uma estrutura formal característica, e por desempenharem no discurso funções normativas, didáticas, moralizadoras, ou outras.

As *frases fixas* (ou *expressões fixas/idiomáticas*), por sua vez, não apresentam a mesma autonomia discursiva que os provérbios, pois são comumente percebidas enquanto enunciados que estão diretamente integrados nos contextos sintáticos das frases em que ocorrem e dependem de elementos desses contextos para a sua completa interpretação. As frases fixas podem ser consideradas *frases elementares* (M. Gross, 1982, 1996; Baptista *et al.* 2015), na medida em que expressam um predicado semântico; são interpretadas globalmente, não podendo ser analisadas a partir da composição do significado das palavras que as constituem quando estas são usadas independentemente. São formadas por um verbo e pelo menos um constituinte que com ele é distribucionalmente fixo ou que só permite a formação de um pequeno paradigma de variação, muito limitado, e distribucionalmente imprevisível. O sujeito destas expressões é, na grande maioria dos casos, um grupo nominal

distribucionalmente livre. Por exemplo, na frase fixa *pôr a carroça à frente dos bois*, o sujeito (humano) é distribucionalmente livre e os dois complementos são fixos com o verbo.

Ainda que os provérbios e as frases fixas constituam expressões de diferente natureza, nem sempre é fácil distingui-los. Isso poderá dever-se ao facto de haver frases fixas que são usadas proverbialmente, e.g. *Não metas o nariz onde não és chamado* (Machado, 1996: 353); *Meter o nariz onde não se é chamado* (Neves, 2000: 279). Esta noção de as frases fixas serem usadas proverbialmente é aqui empregue nas situações em que estas frases fixas (1) são introduzidas por certas expressões formulaicas como, por exemplo, *como diz o provérbio*, *como dizia a minha avó*, etc.; (2) são utilizadas enquanto texto “alheio” (com valor de citação); (3) têm como finalidade advertir, aconselhar, ordenar, instruir, etc.; funcionam como verdades intemporais; ou veiculam normas de conduta.

Pelo contrário, neste caso, além do uso proverbial é possível encontrar a frase fixa com um sujeito livre num contexto narrativo, denotado pelo uso do modo indicativo: “*Raw Kane também meteu o nariz onde não era chamado e resolveu assustar os quatro candidatos ao título quando a conversa ...*” [1]

Por outro lado, há igualmente provérbios que apresentam expressões fixas na sua constituição; e.g. *Quando vires as barbas do vizinho a arder, põe as tuas de molho* (S. Parente, 2005: 564) provérbio que só é interpretável quando comparado com a expressão fixa, de sujeito livre *pôr as barbas de molho* (Neves, 2000: 349): ““*Embora possa não parecer, o sr. ministro pôs as barbas de molho*”, gracejou Cavaco, na resposta.” [2]

De um modo geral, as obras de especialidade (coletâneas ou dicionários), independentemente de tratarem de expressões de natureza paremiológica ou fraseológica, não distinguem claramente estes dois tipos de estruturas, o que também poderá contribuir para essa confusão. Algumas propriedades sintáticas poderão, todavia, ser um indicador do estatuto destas expressões ou como frases fixas, ou como provérbios. O facto de as frases fixas apresentarem normalmente uma ou mais posições distribucionalmente livres (tipicamente, o sujeito) poderá ser um ponto de partida para identificar algumas destas expressões.

Ora, neste sentido, observa-se que muitas expressões recolhidas em coletâneas e dicionários de provérbios apresentam *sujeitos indefinidos*, e.g. *Quem tudo quer tudo perde*, o que poderá dificultar a determinação dos limites entre frases fixas e provérbios. Importa, pois, encontrar critérios claros e operativos (reprodutíveis) que nos permitam determinar as fronteiras entre provérbios e expressões fixas em particular para aquelas expressões em que um *sujeito indefinido* aceita transformações próprias das frases fixas (com sujeito livre), nomeadamente ao nível da variação no preenchimento lexical da posição argumental de sujeito.

Este artigo está organizado do seguinte modo: Num primeiro momento apresenta-se as considerações de alguns autores de recolhas paremiológicas acerca do tipo de materiais presentes nas coletâneas de provérbios (§1); depois é apresentado o que se entende por *sujeito indefinido* (§2); na secção seguinte, expõe-se as características gerais dos provérbios e são determinadas as fronteiras entre provérbios e expressões fixas (§3); na seção 4 é apresentado o *corpus* de trabalho utilizado, bem como a metodologia empregue; depois, na secção seguinte (§5), apresenta-se as principais estruturas proverbiais com sujeito indefinido do nosso estudo, e é validada a existência de expressões proverbiais com sujeito indefinido no discurso, com base em usos reais na *web*; por último, apresenta-se as conclusões e perspectivas de trabalhos futuros (§6).

1. Materiais em recolhas paremiológicas

Várias coletâneas de provérbios apresentam uma grande diversidade de expressões e não apenas proverbiais, tendo os autores dessas obras plena consciência desta variedade, de que, por vezes, dão conta nas suas introduções. Por exemplo, Machado (1996: 7) afirma:

Não pretendo desenvolver considerações sobre os conceitos de adágio, aforismo, anexim, apotegma, axioma, ditado, [...] dito, dizer, exemplo, máxima, parémia, preceito, próloquio, provérbio, refrão, rifão, sentença, etc. [...] Até pode não haver facilidades em os distinguir dos de algumas das chamadas locuções idiomáticas. Na verdade, também não os sei distinguir com exactidão, mas, como creio, não estou só nessa ignorância.

Parente (2005: 5), por sua vez, diz:

[...] muitas vezes temos dificuldade em distinguir com clareza meridiana conceitos muito próximos ou de fronteiras mal definidas – rifão, ditado, provérbio [...] breves sintagmas, locuções, ditos populares...

Já Costa (1999: 11-14) refere:

[...] Mas, logo que se começou a ordenação dos primeiros quatro ou cinco mil ditados recolhidos e houve que buscar fontes menos disponíveis, não foram poucas as dúvidas que me assaltaram e as contradições que encontrei. E isto porque alguns autores faziam distinção, uns de forma mais nítida do que outros, dos nomes dados ao que eu, numa primeira abordagem, entendi ser uma única substância.

[...]

Quis evitar os ditos comuns, as frases feitas, as interjeições compostas, muito do que hoje também chamamos de bocas. Mas, se umas tantas das duas ou três centenas de bocas que constam do longo rol, merecem a companhia dos trinta e um mil vizinhos porque encerram, na verdade, nem sempre de uma forma convencional e bem evidente, um conceito ou máxima, acontece que, algumas outras, de menor merecimento, vieram por arrastamento.

Conscientes destas dificuldades, interessa-nos neste trabalho ter um melhor conhecimento sobre as expressões que apresentam *sujeitos indefinidos*, como as que seguidamente apresentamos:

- (1) *Cada um puxa a brasa à sua sardinha*
- (2) *Albarde-se o burro à vontade do dono*
- (3) *Ninguém deixe o certo pelo duvidoso*
- (4) *Todos têm a sua cruz*
- (5) *Muitos entram lambendo e saem mordendo*
- (6) *Quem semeia ventos colhe tempestades*

Os exemplos (1)-(6) apresentam aquilo a que chamamos *sujeitos indefinidos*, seja pelo facto de esta função sintáctica ser desempenhada por um pronome indefinido: *cada* (1), *ninguém* (3), *todo* (4), *muito* (5); seja pelo facto de o sujeito ser uma oração completiva pseudo-interrogativa introduzida por um pronome interrogativo, *quem* (6); seja ainda pelo emprego de pronome indefinido *-se* (2). A fim de melhor definir o conceito de *sujeito indefinido* que aqui usamos, veja-se a secção seguinte.

2. Sujeitos indefinidos

O consenso possível quanto a diferentes tipos de *sujeito indefinido* – enquanto mecanismo linguístico para veicular *vagueza*, corresponde não só aos casos designados na terminologia tradicional por *sujeitos indeterminados* (cf. i-v), como também aos casos descritos seguidamente em vi-ix, e que incluem algumas construções com verbos no infinitivo (pessoal e impessoal) (cf. vi); as construções em que a forma de sujeito é representada por pronomes

indefinidos simples ou compostos (cf. vii e vii); e também as construções com constituintes interrogativos na função de sujeito de orações subordinadas substantivas (interrogativas indiretas) (pseudo-interrogativas), e.g. *quem, o que*.

A fim de verificar se o material presente nas recolhas de provérbios aqui usadas (base de dados) apresentava ou não sujeito distribucionalmente livre, partimos das expressões que apresentam um dos tipos de *sujeito indefinido* e construimo-las com um grupo nominal nessa posição de sujeito. Neste conjunto de construções incluem-se:

- i. as expressões sem sujeito exposto (cf. “sujeito nulo”);
- ii. em que não é possível identificar o antecedente do sujeito omissivo a partir do contexto;
- iii. em que o sujeito pode ser parafraseado por *alguém, as pessoas* (em geral);
- iv. expressões em que o verbo se encontra na 3ª pessoa do plural, não havendo referência anterior a uma entidade que pudesse ser antecedente desse sujeito, pelo que não é possível reconstituí-lo, e.g. *Telefonaram-me mas não atendi. / Roubaram-me a carteira no cinema;*
- v. expressões com o verbo na 3ª pessoa do singular construído com pronome indefinido *se*, e.g. *Vende-se casas;*
- vi. expressões em que o verbo se encontra no infinitivo impessoal como, por exemplo, as construções imperativas que encontramos em receitas culinárias, e.g. **Juntar** *as gemas com o açúcar, bater muito bem...* ; ou quando o verbo se encontra no infinitivo impessoal com sentido genérico em que não é possível determinar um antecedente para o sujeito: e.g. *É importante [-] comer legumes;* ou quando o verbo se encontra no infinitivo pessoal na primeira pessoa do plural, com sentido genérico, em que não é possível determinar um antecedente para o sujeito: e.g. *É importante [-] comermos legumes;*
- vii. expressões em que a forma de sujeito é representada por *pronomes indefinidos simples* (variáveis ou invariáveis), e.g. *algum, nenhum, todo, outro, muito, pouco, certo, vários, tanto, quanto, qualquer, alguém, ninguém, tudo, outrem, nada, quem, cada, algo;*
- viii. expressões em que a forma de sujeito é representada por *pronomes indefinidos compostos* (e.g. *cada um, cada qual, qualquer um, seja qual for, seja quem for, todo aquele que, etc.*);
- ix. expressões com constituintes interrogativos na função de sujeito de orações subordinadas substantivas (interrogativas indiretas) *quem, o que* (e.g. *Quem sujou, que limpe, O que comprei tem de ir já para o frigorífico*); e as locuções pronominais indefinidas *quem quer*

que, o que quer que, qualquer que, etc. (e.g. Quem quer que tenha tomado essa atitude fê-lo sem pensar);

Na próxima secção apresentamos algumas das características dos provérbios, bem como os princípios que nos regemos para considerar que uma dada expressão não é um provérbio.

3. Características gerais dos provérbios e determinação de fronteiras entre provérbios e expressões fixas

Algumas características formais muito gerais associadas aos provérbios permitem-nos identificá-los de forma relativamente clara. Trata-se de, nomeadamente, dos casos em que o provérbio apresenta uma estrutura bipartida ou bimembre (e.g. *Barriga cheia, pé dormente*); paralelismo de diferentes tipos (e.g. *Quem vê caras não vê corações*); rimas e ritmo (e.g. *Lua de mel, lua de fel*), etc. Sendo estruturas marcadas pela atemporalidade (dado veicular uma verdade universal, exprimem uma máxima, etc.), os verbos dos provérbios encontram-se maioritariamente nos tempos verbais do presente e do futuro, e.g. *Quem boa cama fizer, nela se deitará*. No caso das máximas e.g. *Faz o bem sem olhar a quem*, estas apresentam geralmente os verbos no modo imperativo ou no infinitivo com esse valor imperativo, e.g. *Fazer o bem sem olhar a quem*. Do ponto de vista semântico, são estruturas com um sentido genérico e estão, muitas vezes, associados a uma linguagem metafórica ou figurada. Preferimos, contudo, não tratar como provérbios algumas expressões mesmo que apresentassem algumas destas características.

Neste sentido, consideramos que **não** são provérbios as estruturas que:

- i. apresentam sujeito livre, por exemplo, expressões como:

Perca-se tudo menos a honra [3],

pois a posição sintática de sujeito pode ser preenchida, por exemplo, por um nome humano:

O (Pedro + partido + Eurico) perdeu tudo menos a honra.

- ii. permitem a sua *atualização* no discurso, no sentido que é dado a este termo por G. Gross (1996), ou seja, expressões que podem ser integradas no discurso, por exemplo aceitando variação livre dos tempos verbais.

Les divers aspects de l'actualisation des phrases simples et les éléments lexicaux qui les représentent sont nombreux. On ne retiendra ici, pour le moment, que les informations de temps, de personne et de nombre pour les prédicats et les déterminants pour les arguments. L'actualisation des prédicats dépend de leur catégorie morphologique: désinences verbales ou verbes auxiliaires pour les verbes, verbes supports pour les noms et support être pour les adjectifs.

O facto de os provérbios terem um valor geral de citação e surgirem como um texto encaixado noutra texto, faz com que se suspendam os processos de coesão discursiva, como as relações de correferência ou o encadeamento de tempos-modos verbais; pelo contrário, no caso das expressões fixas, estas estabelecem com o texto relações de correferência (argumentos livres com os seus antecedentes) e o tempo-modo verbal é atualizável (variável). Veja-se, a propósito, o exemplo seguinte que ilustra quer a situação de sujeito distribucionalmente livre, quer a possibilidade de atualização desta expressão no discurso:

“Um Eurico *que perdeu tudo menos a honra, por não quebrar o seu juramento com o Mais Velho.*” [4]

Neste exemplo, o sujeito é distribucionalmente livre (*Um Eurico*) e o verbo *perder* encontra-se no tempo e modo (pretérito perfeito) que é o adequado no encadeamento temporal de toda a frase, isto é, na sequência de um modificador relativo de *Um Eurico* cujo verbo (*salvou*) se encontra no mesmo tempo-modo.

Estes dois critérios são suficientes para distinguir a maioria das expressões idiomáticas ou fixas dos provérbios que constituem o foco deste estudo.

No ponto seguinte, será apresentado o *corpus* de trabalho, bem como a metodologia utilizada neste estudo.

4. Corpus e metodologia

Para a análise das expressões fixas com *sujeito indefinido* deste estudo, partimos de um *corpus* (base de dados) de provérbios com mais de 114.000 entradas (Reis & Baptista, 2016a), que teve por base 4 coletâneas de provérbios (Costa, 1999; Machado, 1996; Moreira, 1996; Parente, 2005).

De forma a podermos identificar as expressões com *sujeito indefinido* aqui tratadas usámos uma plataforma para o processamento de *corpora* textuais – o sistema Unitex (Paumier, 2016) e desenvolvemos um método de pesquisa com base em expressões regulares, como as que apresentamos nestes exemplos,

<V:3s>#-#se: *Julga-se* o ladrão pelo seu coração (#2.776 ocorrências)

<ADV> se <V:3s>: *Cá se fazem cá se pagam* (#3.927 ocorrências)

<CONJ> se <V:3s>: *Enquanto se canta* não se assobia (#1.801 ocorrências),

Que permitem encontrar diferentes combinações de *se* com verbos na 3ª pessoa do singular [5].

Para a análise das expressões com *-se* deste estudo, seguimos os critérios já apresentados em Baptista *et al.* (2018) para a classificação de construções pronominais, que iremos agora indicar:

A distinção entre pronome indefinido ou construção passiva pronominal (também chamada com partícula apassivante) faz-se com base na concordância do verbo:

- (a) com o pronome indefinido o verbo só se conjuga na terceira pessoa do singular;
- (b) com a partícula apassivante, o verbo concorda com o sujeito, que pode ser plural;
- (c) os casos ambíguos, (verbo no singular e objeto/sujeito singular) dá-se preferência à construção com pronome indefinido.

Uma vez que após a identificação automática destas expressões não foi possível aplicar neste sistema um processo de desambiguação morfossintática, torna-se necessário verificar manualmente quais os resultados que correspondem a frases fixas e quais são efetivamente provérbios. Observe-se, a propósito, o seguinte excerto de uma destas concordâncias:

beira. LP_SP26699 O rico come e o pobre [alimenta-se](#). LP_SP26700 O rico domina os pobres, e o de r ajuda os atrevidos. LP_SP24080 O amor [aloja-se](#) nas rugas. LP_SP24081 O amor aperfeiçoa as alm mas sofre-se o que é mau. GLP_JPM02560 [Ama-se](#) o liberal para o desfrutar. GLP_JPM02561 Amanhã, a, andarás carreira chã. GLP_JPM02558_1 [Ama-se](#) a traição, aborrece-se o traidor. GLP_JPM02558 A e mantém chamo pai e mãe. LPP_JRMC22762 [Ama-se](#) o bom pai mas sofre-se o que é mau. LPP_JRMC2276 ição, aborrece-se o traidor. LP_SP03924 [Ama-se](#) mais com o coração que com a cabeça ou as mãos. er, só a Deus se concede. LPP_JRMC02085 [Ama-se](#) sem razão e sem razão se odeia. LPP_JRMC02086 Am a vida vã. LP_SP01988 A mulher não ama: [ama-se](#). LP_SP01989 A mulher não pode e não tem o direit MC05862 No casamento, o principal não é [amar-se](#), é conhecer-se. LPP_JRMC05863 No dia em que te e o burro como o dono manda. LP_SP03920 [Amarra-se](#) o burro onde (aonde) o dono manda. LP_SP03921 o onde (aonde) o dono manda. LP_SP03921 [Amarra-se](#) o cavalo à vontade do dono. LP_SP03922 Amas d tar, não há governo popular. LP_SP16023 [Governa-se](#) a boca conforme a bolsa. LP_SP16024 Governa-

Figura 1: *Excertos da concordância do padrão <V:3s>#-#se*

Nas primeiras duas linhas da concordância acima, apresentam-se duas construções reflexas com sujeito exposto, *pobre* e *amor*, nomeadamente, e que não serão analisadas neste estudo. Todas as outras expressões não têm sujeito exposto. Em tais casos, o objetivo é determinar se essas expressões são efetivamente provérbios ou expressões fixas.

No ponto seguinte deste trabalho, iremos apresentar e quantificar os principais tipos de *sujeitos indefinidos* encontrados no *corpus*.

5. Estruturas proverbiais com *sujeito indefinido* – tipos

Estabelecidos os critérios que permitem identificar estruturas com *sujeitos indefinidos* (cf. §3) e determinados os métodos a usar, respetivamente ao uso de expressões regulares (palavras como, por exemplo, *cada qual*, *quem*, *tudo*, etc. e outros padrões de caracteres), foi possível identificar as principais estruturas proverbiais com sujeito indefinido no nosso *corpus* de trabalho. Na tabela seguinte (Tabela 1), estão representados os principais padrões usados e que apresentam um maior número de ocorrências.

A primeira coluna apresenta os padrões morfossintáticos em estudo, a segunda dá exemplos de provérbios desses padrões e a última coluna indica o número total de ocorrências de cada um desses padrões na nossa base de dados.

Tabela 1, Principais padrões usados para identificar estruturas proverbiais com sujeito indefinido

Padrões	Exemplo	Total
quem	Quem cala consente.	19.042
(<E>+<ADV+Neg>)	Antes cegar que mal ver .	18.337
(<V>+<ADV>+<CONJ>)	Felicidade é não carecermos de a ter.	
<V:W>+<V:V1p>		
(<ADV>+<CONJ>) se	Perdoa- se enquanto se ama.	8.504
<V:3s>+<V:3s>#-# se		
<o> que	O que é bom por si se gaba.	5.421
tudo	Com tempo e esperança, tudo se alcança..	2.231
ninguém	Ninguém as faz que as não pague.	1.605
cada <um>	Cada um é como cada qual.	695
muitos	Muitos falam e exortam, e poucos obram.	607
<todo.PRO:p> (<!N>+<E>)	Todos são iguais perante a lei.	465
(<!N>+<E>) <V:3p>		
cada qual	Cada qual com seu igual.	274

Legenda

Notações: Códigos flexionais <V>: verbo, <W>: infinitivo impessoal, <U>: infinitivo pessoal, <s>: singular, <p>: plural, 1,2,3: 1ª, 2ª e 3ª pessoas, <ADV>: advérbio, <N>: nome, <PRO>: pronome, <CONJ>: conjunção e

<ADV+Neg>: advérbio de negação; <E>: representa a palavra vazia *ípsilon*, o carácter ‘!’ representa a negação de um padrão, o carácter ‘+’ representa a união de expressões regulares; os parênteses angulares (chevrons) ‘<>’ podem ainda representar todas as variações morfológicas associadas a um lema, por exemplo, o padrão <o> permite encontrar as palavras *o*, *a*, *os* e *as*. Neste sistema, as formas representadas em minúsculas ignoram a distinção maiúsculas/minúsculas. Assim, o padrão *ninguém* permite encontrar *ninguém* e *Ninguém*.

Como se pode verificar, os padrões mais frequentes, por ordem decrescente de frequência, dizem respeito às construções com *quem* (#19.042) – destas, 12.516 apresentam *Quem* em início de frase; e às construções com o verbo no infinitivo (#18.337) – impessoal <V:W> ou pessoal (1ª pessoa do plural) <V:V1p>. Nestes casos específicos, será necessário desenvolver métodos mais precisos de identificação das expressões-alvo, de forma a que os resultados sejam mais refinados, e facilitando assim o processo de identificação manual das expressões fixas neste *corpus* de provérbios. Dado o excessivo número de ocorrências encontradas, deixámos a análise destas para outro momento.

Todos os restantes padrões com frequência inferior a 9K já poderão, à partida, ser analisados manualmente.

Foram ainda usados outros padrões, mas, dado o escasso número de resultados obtidos, não foram representados na tabela acima.

Relativamente às construções sem sujeito expresso e verbo na 3ª pessoa do plural, seria necessário dispor de informação sintática, o que neste momento não é possível neste sistema, nomeadamente, capturar os verbos para os quais não tivesse sido capturada uma dependência de sujeito.

Tendo em consideração alguns dos obstáculos encontrados no processo de identificação de expressões com *sujeito indefinido*, procedentes da desambiguação morfossintática e de indisponibilidade de informação sintática, e também pelo facto de se tratar de um trabalho demorado e minucioso, não foram analisados todos os resultados obtidos nesta fase do trabalho.

Dadas todas estas limitações, limitámo-nos, neste estudo, a analisar alguns destes padrões, cujos resultados apresentamos na Tabela 2, nomeadamente, os padrões (<ADV>+<CONJ>) *se* <V:3s>+<V:3s>#-#*se* e <todo.PRO:p> (<!N>+<E>) (<!N>+<E>) <V:3p> (simplificados na Tabela 2), *ninguém*, *muitos*, *cada* <um> e *cada qual*.

Tabela 2, Análise das estruturas proverbiais com sujeito indefinido

Padrão	nº total de ocorrências	nº de provérbios	%	nº de EFs	%
<V:3s>#-#se [6]	8.504	8.241	97%	263	3%
ninguém	1.605	1.554	97%	51	3%
cada <um>	695	548	79%	147	21%
muitos	607	600	99%	7	1%
<todo.PRO:p> <V:3p>	465	462	99%	3	1%
cada qual	274	157	57%	117	43%

No caso dos padrões <V:3s>#-#se e **ninguém**, apenas 3% correspondiam a expressões fixas, isto é, com sujeito distribucionalmente livre, usadas proverbialmente, e.g.

Provérbio: LP_SP23466 *No dia de S. Martinho, **come-se** castanhas e **bebe-se** vinho*

EF: GLP_JPM12988 *Mata-se a cobra e **mostra-se** o pau.*

“O Adão *matou a cobra e mostrou o pau, então ninguém mais segurou os dois...*” [7]

Provérbio: LPP_JRMC13103 *Tosse, amor e febre, **ninguém** esconde*

EF: GLP_JPM16018 *Ninguém faz omeletes sem ovos.*

“Um treinador *não faz omeletes sem ovos...*” [8]

No que respeita aos padrões **muitos** e <todo.PRO:p> <V:3p>, somente 1% correspondiam a EFs. Veja-se os seguintes exemplos de provérbios e expressões fixas com esses padrões:

Provérbio: LPP_JRMC14993 *Muitos morrem na guerra, mas mais vão a ela*

EF: GLP_JPM13737 *Muitos vão buscar lã e voltam tosquiados.*

“Augusto Inácio *foi buscar lã e saiu tosquiado. Imperador é imperador.*” [9]

Provérbio: LPP_JRMC03076 *A árvore caída, **todos** vão buscar lenha*

EF: LPP_JRMC16172 *Todos têm a sua hora*

“António Vitorino, perdão, queria dizer o humorismo também **tem a sua hora** e nem sempre o Sr. Deputado tem a graça que costuma ter” [10]

Quando comparamos a percentagem de *EFs* por número de ocorrências capturadas pelos padrões com os pronomes indefinidos *cada um* e *cada qual*, verifica-se para este último que essa percentagem é cerca de o dobro da verificada para o primeiro (43% e 21%, respetivamente), e.g.

Provérbio: LPP_JRMC07369 *Cada qual folgue com seu igual*

EF: LP_SP07402 *Cada qual aprende à sua custa*

“Contudo, tal como o PSD **aprendeu à sua custa**, quem governa fica com os créditos e quem está na oposição só convence os convencidos.” [11]

Provérbio: GLP_JPM26200 *Tanto vale cada um na praça, quanto vale o que tem na caixa*

EF: GLP_JPM05168 *Cada um julga os outros por si*

“O sampaiense **julga os outros por si**. Mas daqui a alguns meses vai flar de outra maneira. Que eu saiba a Junta não cai nem cairia, as pessoas é que não têm palavra.” [12]

Por outro lado, verifica-se que alguns provérbios admitem a variação livre de *cada um* com *cada qual*:

(1) GLP_JPM05119 *Cada um puxa a brasa à sua sardinha*

(1a) LPP_JRMC17005 *Cada qual puxa a brasa à sua sardinha*

(2) GLP_JPM05146 *Cada um é para o que nasce*

(2a) LPP_JRMC27277 *Cada qual é para o que nasce*

(3) LP_SP07508 *Cada um com a sua cruz*

(3a) LP_SP07406 *Cada qual com a sua cruz*

Observe-se também um exemplo em que os membros do provérbio comutam:

(4) LPP_JRMC16683 *Cada um é como cada qual*

(4a) LP_SP07424 *Cada qual é como cada um*

A forma *cada qual* parece estar já a cair em desuso, no português europeu contemporâneo, sendo a sua prevalência nos provérbios talvez um sinal do carácter arcaizante da expressão (Nóbrega & Miyagawa, 2015).

A fim de validar a análise realizada verificámos algumas ocorrências na *web* de expressões tidas como ‘proverbiais’ e incluídas nas coletâneas de provérbios que nos serviram de base [13]. Nestes exemplos é possível não só verificar variação no preenchimento lexical da posição argumental de sujeito, como também a atualização no discurso.

i. *se* indefinido

GLP_JPM02087 *A vingança serve-se fria.*

(1a) “*Poucos dias depois, a formação de Rui Pereira serviu a vingança fria, com os golos da festa a serem apontados por Ricardo Ferreira (16m), Gabriel (19 e 33m), Emerson (36m) e Nandinho (37m).*” [14]

(1b) “*Marcelo passou rapidamente a cenógrafo-pirómano, Santana – como seria de esperar – é avesso a pactos, Menezes serve a vingança fria e sem requinte.*” [15]

(1c) “*Vociferou contra o pai o caminho todo para casa. Diz que para o ano **lhe serve a vingança fria***” [16]

(1d) “*Como **se serve a vingança? Fria, quente ou simplesmente não se serve? Para mim... serve-se: sempre (que possível e sã). E sem temperatura... isso são detalhes.***” [17]

Este primeiro conjunto de exemplos diz respeito à estrutura com *se* e verbo na 3ª pessoa do singular. O pseudo-provérbio *A vingança serve-se fria* é, afinal, uma *EF* atualizável no discurso e com a posição sintática de sujeito livre, aqui ocupada quer pelo constituinte *A formação de Rui Pereira* quer por *Menezes*. Pode também tratar-se de um sujeito omissivo que retoma anaforicamente um elemento anteriormente referido no discurso, como se pode observar no exemplo (1c). O exemplo (1d) mostra uma variação criativa da mesma expressão,

transformada numa frase interrogativa em que o predicativo do sujeito serve de pivô da interrogativa.

ii. Pronomes indefinidos

LP_SP38965 *Todos calçam pela mesma medida.*

(1a) “Os governos *calçam pela mesma medida*; ser bom funcionario publico é ver e ... calar.”
in *Semana Ilustrada (RJ) - 1861 a 1875 Ano 1861/Edição 00046 – Pag: 6* [18]

GLP_JPM05190 *Cada um sabe as linhas com que se cose.*

(2a) “O FC Porto *sabe as linhas com que se cose*” [19]

LPP_JRMC07669 *Quem semeia ventos colhe tempestades.*

(3a) “O vereador laranja *semeou ventos e quase que colhia tempestades* porque o socialista Alberto Mesquita, um antigo praticante de rãguebi, não gostou e perdeu a paciência.” [20]

(3b) “O governo anti-popular do SYRIZA-ANEL *“semeou ventos e colherá tempestades”*” [21]

LP_SP39542 *Tudo tem o seu preço.*

(4a) “O sucesso *tem o seu preço*, e parte desse preço é a impossibilidade de manter activos de qualidade a longo prazo, actuem eles dentro ou fora das quatro linhas.” [22]

GLP_JPM15976 *Ninguém deixe o certo pelo duvidoso.*

(5a) “A psicologia muitas vezes *deixou o certo pelo duvidoso* ao procurar a aplicação e em particular a medida antes de compreender os mecanismos formadores e a significação dos fatores medidos.” [23]

PP_AM06249 *Muitos entram lambendo e saem mordendo.*

(6a) “Paulo de Tarso *entrou lambendo e saiu mordendo*” [24]

Este conjunto de exemplos demonstra que estas expressões apresentam a posição de sujeito sintaticamente livre e podem ser atualizadas no discurso.

iii. Infinitivo impessoal

LPP_JRMC03976 *Fazer bem sem olhar a quem.*

(1a) “*Deve ter morrido feliz e de consciência tranquila. **Fez bem sem olhar a quem**, guiou-se pelos princípios mais simples e valorosos que um homem pode ter. Um Homem com H maiúsculo.*” [25]

(1b) “*Eu era muito mais feliz quando **fazia o bem sem olhar a quem**. Quando ajudava as pessoas simplesmente porque essa era a vontade do meu coração.*” [26]

A expressão *Fazer bem sem olhar a quem*, à semelhança das anteriores, tem sujeito distribucionalmente livre, como se pode verificar através dos exemplos apresentados. Em ambos os exemplos, a expressão encontra-se integrada no discurso, aceitando variação livre dos tempos verbais. Em (1a) o sujeito já foi referido num período anterior do discurso (subentendido) e o verbo *fez* encontra-se no tempo adequado de acordo com o encadeamento temporal de toda a frase (*deve ter morrido / fez*); em (1b) o sujeito omissa da oração subordinada com *fazia o bem* é correferente do sujeito da oração principal, *eu*. Os tempos verbais *era / fazia* seguem igualmente um processo de encadeamento.

Como se pôde confirmar, os critérios estabelecidos em §3 (variação no preenchimento lexical da posição argumental de sujeito e atualização no discurso) permitiram-nos estabelecer fronteiras claras e reprodutíveis entre provérbios e expressões fixas.

Em contrapartida, retivemos como expressões proverbiais os casos em que, apesar da presença do emprego de *sujeito indefinido*, não se observou variação livre de sujeito nem processos de atualização da expressão no discurso, a qual é sempre usada tal qual nos textos. Observe-se os seguintes exemplos:

PP_AM04117 *Em casa onde não há pão, todos ralham e ninguém tem razão.*

(1a) “*complicado (casa q não ha pão tds **ralham e ninguém tem razão**). Por isso nem tanto a terra nem tanto ao mar.*” [27]

(1b) “*A história repete-se: em casa onde não há pão, todas **ralham e ninguém tem razão**.*” [28]

(1c) “*Podemos aplicar o ditado popular, "em casa onde não há pão todas **ralham e ninguém tem razão**".*” [29]

Os exemplos de (1a)-(1c) são alguns dos 8 resultados encontrados ao efetuarmos uma *pesquisa avançada* no motor de pesquisa Google [30], usando os critérios (1) *exatamente esta palavra ou expressão*: “ralham e ninguém tem razão” (2) nenhuma destas palavras: “Todos”, e (3) *site ou domínio pt*. Como se pode observar, todas as ocorrências encontradas são efetivamente provérbios, cuja posição argumental de sujeito é ocupada por *tds* e *todas*. Encontrámos, todavia, a expressão *Citações Ralham E Ninguém Tem Razão* [31] enquanto título de um *site* de citações, o que aparenta ser uma reutilização criativa do provérbio. Ao

usarmos os mesmos critérios de pesquisa, mas substituindo "ralham e ninguém tem razão" por "* e ninguém tem razão", já que o carácter '*' admite qualquer palavra nessa posição, encontramos 15 ocorrências, incluindo algumas expressões livres. Todas as ocorrências que correspondem a provérbios são, no entanto, variantes de uma mesma *unidade paremiológica*. Veja-se os exemplos:

(1d) “*Como se dizia antigamente casa aonde não há dinheiro toda a gente **ralha e ninguém tem razão***” [32]

(1e) “*o Japão parece um lar de reformados, a China uma casa de correção e a Europa uma casa onde não há pão, porque toda a gente **ralha e ninguém tem razão***” [33]

(1f) “*“ONDE NÃO HÁ PÃO, TODO GRITAM E NINGUÉM TEM RAZÃO”*” [34]

Nestes dois primeiros casos, a posição argumental de sujeito é ocupada por *toda a gente*, tratando-se, igualmente, de um sujeito indefinido. A expressão (1f), por sua vez, apresenta variação verbal, já que o verbo utilizado é agora *gritam*.

Usando os mesmos critérios de pesquisa, mas agora relativos ao provérbio *Quem tudo quer tudo perde* ("tudo quer tudo perde" -"quem" site:pt), obtiveram-se 6 ocorrências. Note-se, no entanto, que aparentemente correspondem a erros de digitação/formatação, ou outros. Veja-se:

GLP_JPM24373 Quem *tudo quer tudo perde*.

(1a) “*200 1 ^aQuem **tudo quer, tudo perde**^fAna Oom^gil. Marta Torrão^gnarração de Bárbara Guimarães^gmúsica de Gonçalo Pratas*” [35]

(1b) “*ISSO CHAMA-SE PRUDÊNCIA QEM **TUDO QUER TUDO PERDE** OS MERCADOS INTERNACIONAIS ESTÃO DE PERNAS PARA O AR*” [36]

Estes últimos resultados realçam a fixidez que este provérbio apresenta. Nas coletâneas que serviram se base a este estudo encontram-se as variantes *Quem quer tudo tudo perde*, *Quem muito quer tudo perde*, *Quem tudo perde tudo quer*, não apresentando variação da posição ocupada pelo sujeito, nem variação do tempo-modo verbal.

Observe-se um exemplo com o verbo no infinitivo impessoal:

GLP_JPM02965 *Antes calar que mal falar*.

Neste caso, procurou-se o padrão “*que mal falar*” -“*calar*” site:pt. Apesar das 9 ocorrências encontradas, nenhuma corresponde ao provérbio *Antes calar que mal falar*, sendo todas construções livres, que parcialmente emparelham com o padrão pesquisado, e.g.

(1a) “*Há uns dias a Agnes Arabela, que mal falar sabe, foi para a sua página do facebook criticar a comunidade homossexual e afins*” [37]

Vejam-se agora este exemplo de uma construção com *-se*:

PP_AM02900 **Conhece-se** o marinheiro quando vem a tempestade.

(1a) “*conheces o marinheiro, quando vem a tempestade:prov:PT*” [38]

Este último exemplo diz respeito à única ocorrência encontrada pelo padrão "o marinheiro quando vem a tempestade" -"conhece-se" site:pt:. O verbo *conheces*, encontra-se aqui no presente do indicativo. A variação morfossintática nos provérbios é mais usual em algumas classes de palavras do que noutras. Segundo Chacoto (1994:174) os nomes, os advérbios e os verbos são as classes mais influenciadas por este fenómeno. Os determinantes, as preposições e as conjunções são as menos afetadas. O exemplo acima apresentado encontra-se numa compilação de provérbios na *web*, não estando integrada no discurso.

Conclusões e trabalho futuro

Neste trabalho, foi apresentado o conceito de *sujeito indefinido* e, com base no mesmo, foram construídos padrões de pesquisa que permitem extrair automaticamente expressões proverbiais com sujeitos indefinidos presentes em uma base de dados de frases proverbiais, com mais de 114.000 entradas, recolhidas a partir de 4 coletâneas de provérbios, e ainda quantificá-las. Após a identificação e quantificação destas estruturas, procedeu-se à verificação manual dos resultados obtidos, de forma a poder identificar expressões fixas no *corpus* de trabalho, de acordo com os critérios estabelecidos. Verificou-se que um elevado número de expressões que se encontram nas coletâneas de provérbios e que se assemelham a provérbios são efetivamente expressões fixas. Os critérios estabelecidos permitiram-nos estabelecer fronteiras claras e reprodutíveis entre provérbios e expressões fixas.

Verificou-se também que algumas expressões com sujeito indefinido admitem a comutação com outros elementos indefinidos.

Usando o motor de pesquisa Google, confirmou-se que algumas expressões, embora tradicionalmente classificadas como provérbios, constituem efetivamente expressões fixas.

Num trabalho futuro, pretendemos analisar os restantes resultados já obtidos através dos padrões de pesquisa desenvolvidos, bem como estabelecer novos padrões, mais precisos, para identificar as expressões com o maior número de ocorrências.

Este estudo poderá permitir ainda que sejam estabelecidos critérios mais claros para a seleção do tipo de estruturas a integrar nas coletâneas de provérbios.

Esta investigação foi parcialmente suportada por fundos nacionais através da Fundação para a Ciência e a Tecnologia (ref. UID/CEC/50021/2019).

Notas

1. <https://specialwrestling.blogs.sapo.pt/tag/pintomatch> (data do último acesso: abril de 2019. Todos os restantes URL foram confirmados nesta data.)
2. <https://www.dn.pt/politica/interior/cavaco-sublinha-importancia-de-ligacoes-ferroviarias-rapidas-2550456.html>
3. O pronome funciona como índice de indeterminação do sujeito.
4. <https://aspartesdotodo.blogspot.com/2009/10/old.html>
5. Nestes dois últimos exemplos, o contexto à esquerda do pronome *se* permite delimitar melhor as expressões-alvo a capturar. A utilização de três termos na pesquisa (e não apenas o pronome *se* e a forma verbal) permite, por sua vez, refinar os resultados obtidos, nomeadamente, evitar a captura de algumas construções pronominais (e.g. *Se a velha se abaixa, para si caga.*)
6. Considerou-se também os casos de próclise quando precedidos de um advérbio ou de uma conjunção.
7. <http://cabanodatabauera.blogspot.com/2011/04/ta-biblia.html>
8. <https://blogvisaodemercado.pt/2014/12/ac-milan-encaixa-85-milhoes-de-euros/>
9. <https://www.cmjornal.pt/opiniao/columnistas/leonor-pinhao/detalhe/ir-buscar-la-sair-tosquiado>
10. <http://debates.parlamento.pt/catalogo/r3/dar/01/02/02/064/1982-03-16/2773>
11. <https://www.dn.pt/opiniao/opiniao-dn/nuno-garoupa/interior/estado-das-unioes-9099571.html>
12. <https://rupturavizela.blogs.sapo.pt/703831.html>
13. Nestes exemplos, assinalámos a negrito os elementos da expressão fixa e em contra itálico o sujeito livre.
14. <https://maisfutebol.iol.pt/outros-desportos/futsal-benfica-modicus-freixieiro-olivais-taca/1333165-1493.html>
15. <https://blogdeurgencia.wordpress.com/2008/08/31/a-quem-perten-1/>
16. <https://contosdameninamulher.blogs.sapo.pt/natal-na-amadorah-pt-ii-174459>
17. <https://amulhercerta.blogs.sapo.pt/181840.html>
18. <http://memoria.bn.br/DocReader/Hotpage/HotpageBN.aspx?bib=702951&pagfis=366&url=http://memoria.bn.br/docreader#>
19. <https://www.cmjornal.pt/mais-cm/domingo/detalhe/toni-fcp-sabe-as-linhas-com-que-se-cose>
20. <https://omirante.pt/semanario/2004-01-15/o-cartoon-da-noticia/2004-01-14-dia-santo-na-loja>
21. <http://es.kke.gr/pt/articles/O-governo-anti-popular-do-SYRIZA-ANEL-semeou-ventos-e-colhera-tempestades/>
22. <https://camaroteleonino.blogs.sapo.pt/o-preco-do-sucesso-865404>
23. <https://www.passeidireto.com/arquivo/29383374/livro-psicologia-e-pedagogia-de-jean-piaget/25>
24. <https://robsonpiresxerife.com/notas/paulo-de-tarso-admite-excesso/>
25. <https://jornaldenisa.blogspot.com/2016/11/memoria-medico-amieirense-homenageado.html>
26. <https://www.wattpad.com/501695998-relatos-de-uma-escritora-amadora-fazer-o-bem-sem>
27. <http://www.radioelmo.com/forum/message.asp?IdMsg=2407&idForum=1>

28. <https://www.jn.pt/opiniao/paulo-ferreira/interior/crise-sem-culpados-1015615.html>
29. <http://www.ensino.eu/ensino-magazine/dezembro-2014/politecnico/joaquim-mourato-lan%C3%A7a-debate.aspx>
30. De forma a obtermos resultados o mais fidedignos possível, percorremos até à última página em cada uma das pesquisas efetuadas, o que permite identificar os resultados mais relevantes, omitindo as entradas muito semelhantes às apresentadas. Nestas pesquisas tivemos em consideração quer a expressão *ipsis verbis*, quer as diferentes variantes desses mesmos provérbios, de modo a descartar a hipótese de a não ocorrência de uma expressão poder dever-se à escolha da variante.
31. <https://pt.jarofquotes.com/perfil/ralham%20e%20ningu%C3%A9m%20tem%20raz%C3%A3o/?page=4>
32. <https://biclaranja.blogs.sapo.pt/417476.html>
33. <https://jornaleconomico.sapo.pt/noticias/o-ano-de-aristoteles-258078>
34. <https://www.jornaldasautarquias.pt/noticias-comentarios.php?id=889&pn=12>
35. http://biblioteca.cm-sbras.pt/Opac/Pages/Document/DocumentFormat.aspx?DocsListId=22085&MFN=22085&DataBase=I0073_BMSBA
36. <http://www.destak.pt/artigo/349720>
37. <http://zetugablog.blogspot.com/2015/12/claudio-ramos-estoura-com-parva-maluca.html>
38. <https://natura.di.uminho.pt/~jj/pln/proverbio.dic>

Referências

- BAPTISTA, Jorge, MARQUES, José, FERNANDES, Graça, GUERRA, Joaquim, MARIANO, Alexandra MARQUES, João, NORONHA Fátima & REIS, Sónia, 2018. Emprego dos pronomes clíticos do português europeu em *corpora* de aprendizagem de português língua estrangeira, VII SIELP - Simpósio Internacional de Ensino da Língua Portuguesa, U. Minho, Braga, 16-18 julho 2018 (2018, aceite para publicação).
- BAPTISTA, Jorge, 1994. Estabelecimento e Formalização de Classes de Nomes Compostos, Dissertação de Mestrado, Lisboa: Faculdade de Letras da Universidade de Lisboa.
- BAPTISTA, Jorge, FERNANDES, Graça & CORREIA, Anabela, 2005. “Léxico-gramática das frases fixas do Português europeu. Breve presentatió”. *Cadernos de Fraseología Galega*, Xunta de Galicia, vol. 7, Galicia, Spain, 41–53.
- BAPTISTA, Jorge, FERNANDES, Graça, TALHADAS, Rui, DIAS, Francisco & MAMEDE, Nuno, 2016. “Implementing European Portuguese Verbal Idioms”. In Gloria Corpas Pastor (ed) *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives*. Geneva: Editions Tradulex, 102–115.
- CARNEIRO, Amanda, 2016. Descrição e classificação das expressões cristalizadas com *ser* e *estar* do português do Brasil. Dissertação de Mestrado. Universidade Federal de São Carlos.
- 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.
- COSTA, José, 1999. *O Livro dos Provérbios Portugueses*. Lisboa: Editorial Presença.
- GARCÍA-PAGE SÁNCHEZ, M., 2008. *Introducción a la fraseología española. Estudio de las locuciones*. Barcelona: Anthropos.
- GROSS, Gaston, 1996. “Prédicats nominaux et compatibilité aspectuelle”. In: *Langages*, 30^e année, n^o121, 1996. Les supports, 54–72.
- GROSS, Maurice, 1982. “Une classification des phrases ‘figées’ du français”. In *Revue Québécoise de Linguistique* 11/2, 1982, UQAM, Montréal, 151–185.
- 1996. “Lexicon Grammar”. In Brown, K. & Miller, J. (eds) *Concise Encyclopaedia of Syntactic Theories*. Cambridge: Pergamon, 244–258.

- MACHADO, José, 1996. *O Grande Livro dos Provérbios*, 1ª edição, Lisboa: Editorial Notícias.
- MATEUS *et al.*, 2003. *Gramática da Língua Portuguesa*. 5ª edição revista e aumentada. Lisboa: Caminho.
- MOREIRA, António, 1996. *Provérbios Portugueses*. Lisboa, Editorial Notícias.
- NEVES, Orlando, 1999. *Dicionário de Expressões Correntes*, 2ª ed., Lisboa: Editorial Notícias.
- NÓBREGA, Vitor, & MIYAGAWA, Shigeru, 2015. The precedence of syntax in the rapid emergence of human language in evolution as defined by the integration hypothesis. *Frontiers in Psychology*, 6, 271.
- PALMA, Cristina, 2009. Expressões Fixas Adverbiais Descrição léxico-sintáctica e subsídios para um estudo contrastivo Português-Espanhol. Dissertação de Mestrado, Universidade do Algarve – FCHS, Faro.
- PARENTE, Salvador, 2005. *O Livro dos Provérbios*. 1ª edição, Lisboa: Editora Âncora.
- RANCHHOD, Elisabete, 2003. “O Lugar das Expressões ‘Fixas’ na Gramática do Português”. In Ivo Castro & Inês Duarte (orgs.) *Razões e Emoção*, vol II. Lisboa: Colibri, 239–254.
- REIS, Sónia, 2014. A Correspondência entre Provérbios e Expressões Fixas no Português Europeu. Dissertação de Mestrado, Universidade do Algarve – FCHS, Faro.
- REIS, Sónia. & BAPTISTA, Jorge, 2016a. 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, Universidade do Algarve, Faro, Portugal, 427–446.
- 2016b. “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.
- STEYER, Kathrin, 2015. “Proverbs from a Corpus Linguistic Point of View”. In Hrisztova-Gotthardt, H. & Varga, M. (eds) *Introduction to Paremiology: A Comprehensive Guide to Proverb Studies*. Berlin: De Gruyter, 206–228.

Curriculum vitae

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 é doutor em Linguística, na especialidade de Sintaxe, e professor associado com ‘tenure’ 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 recursos linguísticos e de aplicações computacionais para o ensino da língua.

Email: reis.soniamm@gmail.com, jbaptis@ualg.pt

Capítulo 8

Descrição e classificação sintática das expressões proverbiais do português europeu

Artigo 8

Publicado em: *Diacrítica* 33(2), Revista do Centro de Estudos Humanísticos, pp. 324-334

DOI: doi.org/10.21814/diacritica.412

DESCRIÇÃO E CLASSIFICAÇÃO SINTÁTICA DAS EXPRESSÕES PROVERBIAIS DO PORTUGUÊS EUROPEU

DESCRIPTION AND SYNTACTIC CLASSIFICATION OF EUROPEAN PORTUGUESE PROVERBS

Sónia Reis*

reis.soniamm@gmail.com

Jorge Baptista**

jbaptis@ualg.pt

Os provérbios são expressões de uso generalizado, utilizados em diferentes situações conversacionais e assumindo diferentes funções no discurso em que se integram. Do ponto de vista sintático, este tipo de expressões apresenta uma grande variedade de estruturas. Tendo isto em conta, o objetivo principal deste trabalho é estabelecer uma classificação formal sintática dos provérbios do português europeu. Para tal, pretendemos desenvolver e aprofundar a tipologia de classificação formal proposta por Rassi *et al.* (2014). Por conseguinte, será considerada uma subclassificação para as classes que apresentam um elevado número de provérbios e uma eventual reclassificação de alguns dos tipos. A proposta de classificação foi validada pela anotação independente por dois linguistas de uma lista de provérbios muito usuais, medindo-se depois o acordo entre anotadores, que foi muito elevado. Esta classificação, por sua vez, será o ponto de partida para o desenvolvimento de um procedimento de classificação automática deste tipo de estruturas, e contribuir assim para a elaboração de recursos para diferentes aplicações em Processamento da Linguagem Natural (PLN).

Palavras-chave: Provérbios. Classificação sintática. Português europeu. PLN.

Proverbs are expressions of widespread use, appearing in different conversational situations, and assume different functions in the discourse in which they are integrated. From a syntactic point of view, this type of expression presents a great variety of structures. With this in mind, the main goal of this study is to determine a formal syntactic classification of the European Portuguese proverbs. For this, we intend to develop and extend the typology of the formal classification proposed by Rassi *et al.* (2014). Therefore, a subclassification for the classes with a large number of proverbs, and a possible reclassification of some of the types will be considered. The classification proposal was validated by having two linguists independently annotating a list of very usual proverbs, and then calculating the inter-annotator agreement, which was found to be very high. This classification, in turn, will be the starting point for the development of an automatic classification procedure, and thus contribute to the preparation of resources for different applications in Natural Language Processing (NLP).

Keywords: Proverbs. Syntactic classification. European Portuguese. NLP.

* Universidade do Algarve – FCHS, Campus de Gambelas, Faro, Portugal. 0000-0001-7709-6889:

** Universidade do Algarve – FCHS, Campus de Gambelas, Portugal | INESC-ID Lisboa – L2F, Lisboa, Portugal. 0000-0003-4603-4364:

•

1. Introdução

Os provérbios estabelecem relações discursivas de natureza diferente com o texto à sua volta. Desde os variados mecanismos utilizados para a sua introdução em discurso, mas mesmo na ausência destes, os provérbios são sentidos (e interpretados) como texto alheio, são microtextos citados, trazidos ao discurso a partir do conhecimento partilhado que os falantes têm do mundo e que se encontra cristalizado numa expressão linguística mais ou menos estável.

O emprego deste tipo de estruturas levanta diversos problemas à análise automática de textos. Na medida em que estes surgem como um texto encaixado noutro texto, os processos de coesão discursiva, como as relações de correferência ou o encadeamento de tempos-modos verbais são, muitas vezes, suspensos. Frequentemente, o seu ponto de inserção não é explicitamente assinalado, havendo situações de reutilização/variação criativa dos seus elementos lexicais mais característicos, que implicam o conhecimento prévio das formas de base para a sua completa descodificação. Assim, os provérbios constituem um desafio ao processamento da linguagem natural (PLN), exigindo uma identificação e delimitação mais precisas.

Pelo facto de o uso dos provérbios predominar na oralidade – o que está, talvez, na origem de certas características formais das expressões proverbiais, tais como as marcas mnemónicas de métrica, rima interna, paralelismo, etc.; um dos aspetos mais relevantes para a sua identificação é o tratamento da variação formal (sobretudo lexical) e a união das variantes sob uma única estrutura de base. Neste sentido, já foi desenvolvido um método que permite identificar automaticamente este tipo de expressões em textos, e que consiste na construção manual de transdutores que representam cada unidade paremiológica (provérbios + variantes) (Reis & Baptista 2016a, 2017b, 2017c) e que teve como ponto de partida o trabalho desenvolvido por Rassi *et al.* (2014) para o português do Brasil, que será descrito no ponto 2 deste trabalho, mas que deste difere não só pelo seu escopo/abrangência como também pela metodologia utilizada, uma vez que, se chegou à conclusão de que uma abordagem de geração exclusivamente automática não é suficientemente precisa para os objetivos que se pretendia alcançar.

O passo seguinte será o desenvolvimento de um procedimento de classificação automática deste tipo de estruturas. Esta classificação permitirá reunir expressões formalmente semelhantes de forma a poder aplicar sobre elas métodos de processamento de língua natural em diversas aplicações (*e.g.* identificação de variantes, reconhecimento de usos ‘criativos’ de provérbios, tradução, entre outros).

Esta classificação será desenvolvida sob o ponto de vista sintático, e partindo da classificação sintática proposta por Rassi *et al.* (2014) para os provérbios do português do Brasil. Tendo-se este propósito em mente, pretende-se agora estabelecer uma classificação formal sintática dos provérbios do português europeu, de forma a que, com base nesta classificação, seja posteriormente desenvolvido um procedimento de

classificação automática de uma base de dados de provérbios. O objetivo de tal classificação é permitir o desenvolvimento de métodos mais eficientes de reconhecimento automático de expressões proverbiais em textos e a sua aplicação em diferentes tarefas de PLN (Processamento da Linguagem Natural).

Este trabalho está estruturado da seguinte forma: a secção 2 apresenta os trabalhos relacionados; a secção 3 expõe e descreve as classes formais dos provérbios do português europeu; segue-se a secção 4 onde é apresentada a validação da taxonomia e os resultados obtidos; finalmente a secção 5 apresenta as conclusões e perspetivas de trabalhos futuros.

2. Trabalhos relacionados

Já existem várias propostas de classificação sintática das frases fixas do português europeu, nomeadamente alguns trabalhos sobre provérbios, porém, não há ainda um trabalho que apresente de forma sistemática uma tipologia deste tipo de expressões. Dos vários estudos sobre provérbios do Português Europeu (PE), poucos são que tratam esses provérbios sobre o ponto de vista sintático. L. Chacoto tem vindo a desenvolver trabalhos neste âmbito, fazendo uma análise léxico-sintática deste tipo de estruturas. Por exemplo, Chacoto (1994) analisa a fixidez e a variação dos provérbios e constata que estas estruturas, apesar da fixidez das suas formas, podem apresentar variação de diferentes tipos e resultante de diferentes fatores. Chacoto (2007) faz, também, um estudo comparativo dos provérbios do português europeu e do espanhol iniciados por *quem/quien*; em Chacoto (2008), a autora trata as estruturas comparativas dos provérbios do PE e, em Chacoto (2010), as estruturas condicionais.

Como já referimos num trabalho anterior (Reis & Baptista 2016b), para o português do Brasil, por sua vez, já foi apresentada uma proposta de classificação formal (sintática) de provérbios (Rassi *et al.* 2014), com base numa coleção de três mil quinhentos e dois provérbios (e suas variantes), organizados em quinhentos e noventa e quatro tipos ou formas de base e recolhidos de vários dicionários de provérbios do português do Brasil. Os autores apresentaram um método, baseado em máquinas de estados finitos (Clark, Fox & Lappin 2010), para identificação automática de provérbios em *corpora* de grandes dimensões, que foi testado no *corpus PLN-Br* constituído por Bruckschein *et al.* (2008), com vinte e nove milhões de *tokens* de texto jornalístico retirado da edição do jornal brasileiro *Folha de São Paulo* (1994–2005). Os autores reportaram uma precisão de 60% a 73%, dependendo sobretudo da classe do provérbio e do grau de completude das variantes lexicais registadas no léxico. O número de provérbios encontrados (e suas variantes), face aos de que o léxico dispunha, é, no entanto, reduzido, o que deverá estar diretamente relacionado com a natureza do *corpus*, uma vez que se trata de texto jornalístico.

Nesse trabalho (Rassi *et al.* 2014), os provérbios são classificados, tendo em conta o número de orações/proposições (ou frases elementares) que apresentam: uma, duas ou três orações, correspondendo às classes principais P1, P2 e P3, respetivamente; e também com base nos seus elementos-chave. De forma muito resumida, esta classificação consiste em:

P1

- Construções impessoais (e.g. *Não há regra sem exceção*) – P1F1;
- Construções com verbo copulativo (e.g. *O amor é cego*) – P1F2;
- Negação obrigatória (e.g. *Burro velho não aprende línguas*) – P1F4;
- Anteposição obrigatória do complemento preposicional (e.g. *Para bom entendedor meia palavra basta*) – P1F5;
- Outros provérbios com uma única oração (e.g. *A união faz a força*) – P1F3.

P2

- Orações comparativas (e.g. *Mais vale pouco que nada*) – P2F1;
- Orações coordenadas (e.g. *A palavra é de prata e o silêncio é de ouro*) – P2F2;
- Proposições coordenadas sem verbo (e.g. *Tal pai, tal filho*) – P2F3;
- Interrogativas-sujeito (e.g. *Quem ri por último, ri melhor*) – P2F4;
- Orações subordinadas (e.g. *Os amigos são muitos quando é grande a abundância*) – P2F5;
- Anteposição obrigatória da oração subordinada (e.g. *Quando um burro fala, os outros baixam as orelhas*) – P2F6.

P3

- Tripla coordenação (e.g. *A laranja de manhã é ouro, à tarde é prata e à noite mata*).

Neste trabalho, partindo da tipologia proposta pelos autores acima citados, elaborar-se-á uma tipologia formal para os provérbios do PE, que descreveremos no ponto seguinte deste trabalho, secção 3. Pretende-se com esta proposta, orientar o desenvolvimento de um procedimento de classificação automática dos provérbios de uma base de dados com cerca de cento e catorze mil expressões proverbiais (Reis & Baptista 2016a), coligidas a partir de 4 coletâneas de provérbios portugueses (Costa 1999; Machado 1996; Moreira 1996; Parente 2005).

3. Classificação formal (sintática) das expressões proverbiais do português europeu

Como já indicado anteriormente, o objetivo que norteia este estudo é o estabelecimento de um quadro de classificação formal (sintática) das expressões proverbiais em português europeu. Neste sentido, pretende-se desenvolver e aprofundar a tipologia de classificação formal, de natureza sintática, proposta por Rassi *et al.*, aplicada a provérbios usuais no português do Brasil, mas agora com uma maior base empírica e revendo alguns critérios ali usados, por questões de coerência/consistência interna e simplificação do procedimento de classificação.

Algumas das dificuldades desta classificação prendem-se com o facto de muitas destas expressões apresentarem características que podem ser associadas a diferentes classes formais, segundo a hierarquia taxonómica que as estrutura. De forma a ultrapassar este problema, ter-se-á por base, neste trabalho, uma classificação assente em critérios estritamente formais, o que auxiliará no processo taxonómico. Efetivamente, o critério de

base para a classificação de Rassi *et al.* (2014), nomeadamente, o número de proposições constitutivas dos provérbios, não foi o critério predominante para a classificação aqui apresentada, por se tratar de um critério que revelou uma reprodutibilidade problemática, já que corresponde a um construto teórico e não a uma propriedade formal, empiricamente evidenciada. Por outro lado, outros critérios como a negação obrigatória não foram considerados num primeiro momento, já que esta propriedade é transversal a frases de tipologias muito diversas.

A classificação sintática dos provérbios do português europeu está dividida em 8 classes de base, tendo algumas destas sido subdivididas, conforme mostraremos adiante. Saliente-se que esta subclassificação é um processo em curso, que na sequência da determinação da extensão dos conjuntos assim constituídos, poderá vir a sofrer atualizações/adaptações. De seguida, apresentamos as classes de base, já estabelecidas:

Tabela 1. Classificação formal dos provérbios do português europeu.

Classe	Estrutura	Definição	Exemplo
P1	$0 V w$	construções impessoais	Há males que vêm por bem.
P2	$V=V_{cop}$	construções predicativas	O amor é cego.
P3	$N_1 N_2$	construções sem verbo	Cada cavadela, sua minhoca.
P4	<i>QueF</i>	orações substantivas	Quem escorrega também cai.
P5	$F_1 \text{ Conj}(\text{subord}) F_2$	orações subordinadas (adverbiais e adjetivas)	Quando a esmola é muita, o pobre desconfia.
P6	$F_1 \text{ Conj}(\text{subord-comp}) F_2$	orações comparativa	Debaixo da manta, tanto vale a escura como a branca.
P7	$F_1 \text{ Conj}(\text{coord}) F_2$	orações coordenadas	A conversa não engasga mas empata.
P8	$N_0 V w$	frases simples	Em abril, queima a velha o carro e o carril.

Fonte: elaborado pelos autores.

Legenda: N_0, N_1, N_2 : grupos nominais com a função, respetivamente, de sujeito, 1º complemento ou 2º complemento; 0 : construção impessoal (sem sujeito); w : sequência não especificada de elementos; V : verbo; V_{cop} : verbo copulativo; *Conj*: conjunção (*subord*: subordinativa, *subord-comp*: subordinativa comparativa, *coord*: coordenativa); *QueF*: oração subordinada substantiva; F_1, F_2 : frase ou oração.

P1: Construções impessoais – Esta classe é composta por construções sem sujeito em que os verbos são conjugados apenas na 3ª pessoa do singular. Constam desta classe o verbo *haver* (na aceção de *existir*), por exemplo, *Há males que vêm por bem* e os verbos que exprimem fenómenos atmosféricos e da natureza, por exemplo, *Em Março chove cada dia um pedaço*.

A proposta de Rassi *et al.* (2014) contempla igualmente uma classe de construções impessoais, que difere da que propomos na medida em que abrange as construções com o verbo *ter* (impessoal) conjugado apenas na 3ª pessoa do singular (na aceção de *existir*, e.g. *Tem muito cacique pra pouco índio*, sendo sinónimo de *haver*, e.g. *Há muito cacique pra pouco índio*), usadas na variante do português do Brasil; e abrange ainda construções com sujeito indefinido, marcadas pelo pronome indefinido *se*, por exemplo, *Devagar se vai ao longe*, que aqui foram integradas noutras classes.

P2: Construções predicativas – Esta classe é constituída por construções que constituem frases simples (com um único predicado) e que apresentam um verbo copulativo (*e.g. ser, estar, ficar, continuar, parecer, permanecer, tornar-se, revelar-se, etc.*) e ligam o sujeito a um outro constituinte, que pode ter como núcleo:

- i) um grupo nominal [P2-1], por exemplo, *Dinheiro é **remédio***;
- ii) um grupo adjetival [P2-2], por exemplo, *Amor de parente é **mais quente***;
- iii) um grupo preposicional [P2-3], por exemplo, *Amigo de mesa não é **de firmeza***;
- iv) ou um grupo adverbial [P2-4], por exemplo, *A boa educação fica **bem** em todo o lado*.

Não foram incluídas nesta classe, por corresponderem à definição de outras classes, as seguintes estruturas/construções:

- a) as estruturas comparativas que apresentam verbos copulativos, por exemplo, *Cada um é **como** cada qual* ou *Melhor é dobrar **que** quebrar*; estas integrarão a classe P6;
- b) as orações subordinadas substantivas interrogativas indiretas, por exemplo, *Feliz é **quem feliz se julga*** ou *O pior cego é o **que não quer ver***; estas integrarão a classe P4;
- c) as orações subordinadas infinitivas, por exemplo, *Partir é **morrer um pouco***; estas integrarão a classe P4;
- d) as orações reduzidas (substantivas) que equivalem a um grupo nominal com função de sujeito, por exemplo, ***Errar** é humano*; estas integrarão a classe P4;
- e) outras frases que integram orações subordinadas (adjetivas e adverbiais), por exemplo, *O sol quando nasce é **para todos***; estas integrarão a classe P5;
- f) outras frases que integram orações coordenadas, por exemplo, *O dinheiro é bom **companheiro, mas mau conselheiro***; estas integrarão a classe P7;
- g) e as construções que apresentam *é que* (muitas vezes considerado um elemento expletivo/enfático), e que não apresentam nenhuma das características já descritas nas alíneas anteriores, por exemplo, *A intenção é **que conta*** (*cf. Não é só a intenção que conta*); de um modo geral, a expressão *é que* não é considerada na classificação.

Foi necessário subdividir esta classe dado o elevado número de provérbios que apresentam um verbo copulativo na sua constituição.

Para o português do Brasil foi igualmente proposta uma classe que abrange as construções com verbo copulativo, a qual contempla os verbos *ser* e *estar*.¹

P3: Construções sem verbo – Esta classe integra as construções que não apresentam verbo (construções elípticas), por exemplo, *Muita parra, pouca uva*. Equivale à classe P2F3 da classificação proposta por Rassi *et al.* (2014).

P4: Orações substantivas – Esta classe inclui as orações subordinadas substantivas completivas e relativas sem antecedente (interrogativas indiretas), por exemplo, *Quem*

¹ A classificação apresentada por Rassi *et al.* (2014) não elenca, porém, nenhum provérbio deste tipo com o verbo copulativo *estar*. Trata-se de expressões como, por exemplo, *No meio é **que está** a virtude*. Encontramos, além disso, provérbios com o verbo *estar* integrados em outras classes, nomeadamente os provérbios *O mal está nos olhos de quem o vê* e *As melhores essências estão nos menores frascos*, que constam da classe P1F3 (*N₀ v w*).

tudo quer tudo perde; O que é doce nunca amargou ou *Onde entra o sol, não entra o médico*, e equivale à classe P2F4 de (*idem: ibidem*), mas incluindo agora as orações introduzidas pelo advérbio *onde*, sem nome ou expressão a que este esteja associado (sem referente).

P5: Orações subordinadas (adverbiais e adjetivas) – Desta classe constam as orações subordinadas adverbiais (não comparativas) e as orações subordinadas adjetivas (relativas). É ainda tido em consideração se a locução ou conjunção que introduz estas orações se encontra anteposta à oração principal [P5-1], por exemplo, *Quando a esmola é muita o pobre desconfia*; ou depois da oração principal [P5-2], *Mal ladra o cão, quando ladra de medo*; o que determinou o desenvolvimento das subclasses [P5-1] e [P5-2]. No que respeita à subordinação, e considerando a ordem das orações (principal e subordinada), Rassi *et al.* (2014) consideraram esta como critério taxonómico, constituindo duas classes distintas de provérbios (P2F5 e P2F6), que aqui aparecem sob a mesma classe principal (P5).

P6: Orações comparativas – Esta classe integra as orações subordinadas adverbiais comparativas (introduzidas por conjunções ou locuções comparativas), por exemplo, *Voa o tempo como o vento, Mais vale tarde que nunca, Antes dobrar que quebrar* ou *Tal é o Demo como sua mãe*²; esta classe corresponde à classe P2F1 de Rassi *et al.* (2014).

P7: Orações coordenadas – Desta classe constam as orações coordenadas (com verbo explícito); estas podem ser sindéticas, por exemplo, *A justiça tarda mas não falta*; ou assindéticas (sem a conjunção coordenativa expressa); esta classe equivale à classe P2F3 da classificação proposta por Rassi *et al.* (2014).

P8: Frases simples – Esta classe é constituída por todas as frases simples que não tenham sido previamente integradas nas classes definidas anteriormente; estas expressões podem apresentar inversão dos constituintes [P8-1], por exemplo, *Em abril queima a velha o carro e o carril* ou podem atender à ordem básica dos constituintes [P8-2], por exemplo, *Bom madeiro corta-se em janeiro*. Para o português do Brasil foram estabelecidas duas classes distintas para representar estas frases com uma única proposição: P1F5 quando há anteposição de complemento e são iniciadas por preposição; e P1F3 quando apresentam os seus elementos na ordem básica.

De seguida iremos apresentar a metodologia utilizada para validar a taxonomia proposta.

² Ainda que alguns autores incluam nesta classe as orações que expressam conformidade e proporção (nomeadamente Chacoto (2008)), estas não foram aqui incluídas, já que constam da classe P5. Seguimos a proposta de classificação de Mateus *et al.* (2003, p. 762), que consideram que “Ao contrário das orações comparativas, as orações conformativas são deslocáveis (...), podem ser objeto de clivagem (...), e são adjuntos”.

4. Validação da taxonomia: classificação do mínimo paremiológico do português e resultados

Com vista a determinar a reprodutibilidade dos critérios taxonómicos acima enunciados, procedeu-se à classificação sistemática do *mínimo paremiológico* do português europeu. Por *mínimo paremiológico* entende-se o conjunto de provérbios mais conhecidos e frequentemente utilizados de uma comunidade cultural (Permjakov 1973). Neste caso, trata-se de um conjunto de trezentos e dezassete expressões proverbiais, cuja disponibilidade lexical foi atestada de múltiplas formas (com recurso a informantes, dados de *corpora* e frequências observadas em textos na internet) (Reis & Baptista 2017a).

Para tal, dois anotadores, independentemente, procederam à classificação das expressões do *mínimo paremiológico*, utilizando para isso um conjunto inicial de diretivas de anotação em que se explicitavam os critérios e se davam exemplos e contraexemplos. Os anotadores foram dois linguistas, falantes nativos do português europeu, com conhecimento de todos os provérbios alistados, bem como das condições pragmáticas do seu uso.

Realizada a anotação, os dados foram comparados, utilizando-se o software *ReCal 0.1 Alpha para 2 anotadores*³ para calcular a concordância entre anotadores. Das trezentas e dezassete expressões, os 2 anotadores concordaram trezentas e dez vezes (97,8%) e discordaram apenas 7 (Krippendorff alfa = 0.973). Tal representa um grau de concordância muito elevado.

Os casos em que houve discordância foram os seguintes (*vd.* Tabela 2):

Tabela 2. Classificações diferentes.

ID-Prov	Provérbio	Anot 1	Anot 2
MP_P47	Aprender até morrer.	P4	P5
MP_P70	Cada um sabe as linhas com que se cose.	P4	P5
MP_P111	Errar é humano.	P2	P4
MP_P112	Faz o que eu digo e não faças o que eu faço.	P7	P4
MP_P151	Não se pode ter tudo.	P4	P8
MP_P157	Ninguém é de ferro.	P2	P8
MP_P170	O fruto proibido é o mais apetecido.	P6	P2

Fonte: elaborado pelos autores.

Na expressão MP_47, houve confusão do primeiro anotador quanto ao estatuto da oração subordinada *até morrer*; em MP_70, *idem*, relativamente à oração adjetiva relativa *com que se cose*; em MP_111, *idem*, relativamente à oração reduzida (substantiva) *errar*; em MP_112, o primeiro anotador não considerou as orações interrogativas indiretas (critério num patamar hierarquicamente superior), dando prioridade à coordenação; em MP_151, o primeiro anotador, certamente por lapso, selecionou uma classe incorreta, já que *poder* é apenas um verbo auxiliar modal de *ter*, aqui empregue como verbo pleno; em MP_157, o segundo anotador não considerou o verbo copulativo; por último, em MP_170, o

³ Disponível em: <http://dfreelon.org/utills/recalfront/recal2/#doc>.

primeiro anotador escolheu a classe P6, considerando a comparação (sem segundo termo de comparação) em detrimento do verbo copulativo (critério superior).

Após esta fase, as divergências foram resolvidas, e os critérios de classificação “afinados” de modo a serem mais consistentes e explícitos.

Verifica-se, portanto, que os critérios de classificação são suficientemente claros, explícitos e reprodutíveis para poderem ser empregues com segurança na tarefa de classificação das restantes expressões proverbiais da base de dados (com cento e catorze mil expressões) que constitui o nosso *corpus* de trabalho.

Esta listagem do *mínimo paremiológico* com a respetiva classificação sintática será o embrião que levará à constituição de uma coleção dourada (*golden standard*), que permita avaliar métodos de classificação automática de provérbios (trabalho em curso). Esses métodos deverão, então, permitir outros processamentos mais complexos, tais como a identificação de variantes de provérbios sob uma mesma entidade, a unidade paremiológica.

5. Conclusões e trabalho futuro

Neste trabalho foi apresentada uma proposta de classificação formal (sintática) para os provérbios do português europeu. Esta proposta desenvolve e aprofunda a classificação de Rassi *et al.* (2014), mas com uma maior base empírica, e revê alguns desses critérios, a fim de garantir uma maior consistência interna e uma simplificação do procedimento de classificação. Em concreto, abandonámos o critério daqueles autores quanto ao número de proposições, por se tratar de um construto teórico e não de um padrão formal, diretamente observável. Também o critério da negação obrigatória foi preterido nesta fase, já que é transversal a várias classes formais. Além disso, os critérios de classificação foram ordenados para uma aplicação sucessiva, a fim de permitir o desenvolvimento de um algoritmo de classificação, que permitisse no futuro automatizar o processo, usando métodos de processamento de língua natural.

Foram, assim, constituídas 8 classes formais principais, baseadas no tipo de construção sintática que o provérbio exhibe, nomeadamente construções impessoais, predicativas (com verbo copulativo), sem verbo, subordinadas (substantivas), subordinadas (adverbiais e adjetivas), construções comparativas, orações coordenadas e outras frases simples.

A estrutura taxonómica constituída foi avaliada através da classificação sistemática do *mínimo paremiológico*, feita autonomamente por dois anotadores, revelando um grau de concordância bastante elevado (97,8%). As diferenças observadas permitiram melhorar as diretivas de classificação, tornando os critérios mais claros e mais explícitos.

Os critérios de classificação são suficientemente claros, explícitos e reprodutíveis para poderem ser empregues com segurança na tarefa de classificação das restantes expressões proverbiais da base de dados.

No futuro, pretendemos desenvolver procedimentos para a classificação automática de expressões proverbiais com base neste trabalho; e eventualmente métodos que venham complementar os que usamos neste momento para identificar expressões proverbiais em

textos, em especial as que não se encontram ainda recenseadas ou que resultem de uma variação criativa pelos falantes.

Financiamento:

Esta investigação foi parcialmente suportada por fundos nacionais através da Fundação para a Ciência e a Tecnologia (ref. UID/CEC/50021/2019).

Referências

- Bruckschen, M., Muniz F., Souza, J. G. C., Fuchs, J. T., Infante, K., Muniz, M., ... Aluísio, S. (2008). Anotação linguística em XML do Corpus PLN-BR. *Série de Relatórios do NILC, NILC-NILC-TR-08-09*. Universidade de São Paulo, Brasil. Disponível em: http://www.nilc.icmc.usp.br/nilc/index.php/publications#conference_papers
- Chacoto, L. (1994). *Estudo e formalização das propriedades léxico-sintáticas das expressões fixas proverbiais* (Tese de mestrado, Faculdade de Letras da Universidade de Lisboa, Lisboa).
- Chacoto, L. (2007). A sintaxe dos provérbios. As estruturas quem/quien em português e espanhol. *Cadernos de Fraseoloxía Galega*, 9, 31–53.
- Chacoto, L. (2008). Vale mais um gosto na vida que três vinténs na algibeira - Las estructuras comparativas en los provérbios portugueses. In G. Conde Tarrío (Ed.), *Aspectos formales y discursivos de las expresiones fijas* (pp. 87–103). Frankfurt: Peter Lang.
- Chacoto, L. (2010). Não há rifão velho, se é dito a propósito – La condición en los refranes portugueses. In J. Korhonen, W. Mieder, E. Piirainen, & R. Piñel (Eds.), *Actas do Congresso Internacional Europhras 2008* (pp. 58–65). Universidade de Helsínquia, Finlândia.
- Clark, A., Fox, C. & Lappin, S. (Eds.) (2010). *The handbook of computational linguistics and natural language processing*. New Jersey: Wiley-Blackwell.
- Costa, J. (1999). *O Livro dos Provérbios Portugueses*. Lisboa, Portugal: Editorial Presença.
- Machado, J. (1996). *O Grande Livro dos Provérbios* (1ª ed.). Lisboa: Editorial Notícias.
- Mateus, M. H. M., Brito, A. M., Duarte, I., Faria, I. H., Frota, S., Matos, G., ... Villalva, A. (2003). *Gramática da Língua Portuguesa* (5ª ed. revista e aumentada). Lisboa: Caminho.
- Moreira, A. (1996). *Provérbios Portugueses*. Lisboa: Editorial Notícias.
- Parente, S. (2005). *O Livro dos Provérbios* (1ª ed.). Lisboa: Editora Âncora.
- Permjakov, G. L. (1973). On the paremiological level and premiological minimum of language. *Proverbium*, 22, 862–863.
- Rassi, A., Baptista, J. & Vale, O. (2014). Automatic detection of proverbs and their variants. In M. J. Pereira, J. P. Leal & A. Simões (Eds.), *3rd Symposium on Languages, Applications and Technologies (SLATE'14)*, (pp. 235–249). Leibniz: Dagstuhl Publishing. <https://doi.org/10.4230/OASlcs.SLATE.2014.235>
- Reis, S. & Baptista, J. (2016a). 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*, 5-6 maio (pp. 427–446). Faro: Universidade do Algarve. <https://doi.org/10.13140/RG.2.1.1244.7603>
- Reis, S. & Baptista, J. (2016b). Portuguese Proverbs: Types and Variants. In G. C. Pastor (Ed.), *Computerised and corpus-based approaches to phraseology: Monolingual and multilingual perspectives* (pp. 208–2017). Geneve: Editions Tradulex.
- Reis, S. & Baptista, J. (2017a). Estimating lexical availability of European Portuguese proverbs. In R. Mitkov (Ed.), *Lecture Notes in Computer Science: Vol. 10596. Computational and corpus-based phraseology. EUROPHRAS 2017* (pp. 232–244). Cham: Springer. https://doi.org/10.1007/978-3-319-69805-2_17

- Reis, S. & Baptista, J. (2017b). O uso de provérbios no ensino de português. In R. Soares & O. Lauhakangas (Eds.), *10º Colóquio interdisciplinar sobre provérbios. Actas ICP16* (pp. 521–538). Tavira, Portugal: AIP–IAP.
- Reis, S. & Baptista, J. (2017c). Os provérbios em manuais de ensino de Português Língua Não Materna. In V. Pinheiro & G. H. Paetzold (Eds.), *Jornadas de Descrição do Português, integradas no STIL2017*, 2–5 outubro (pp. 247–255). Uberlândia: Sociedade Brasileira de Computação.

[recebido em 1 de junho de 2019 e aceite para publicação em 19 de outubro de 2019]

Capítulo 9

Determinação de um mínimo paremiológico do português europeu

Artigo 9

Publicado em: *Acta Scientiarum. Language and Culture*, 42(2)

DOI: e52114. <https://doi.org/10.4025/actascilangcult.v42i2.52114>



Determinação de um mínimo paremiológico do português europeu

Sónia Reis^{1*} e Jorge Baptista^{1,2}

¹Universidade do Algarve, Estrada da Penha, 139, 8005-139, Faro, Portugal. ²Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento de Lisboa, Lisboa, Portugal. *Autor para correspondência. E-mail. reis.soniamm@gmail.com

RESUMO. O principal objetivo deste estudo é apresentar o ‘mínimo paremiológico’ do português europeu, isto é, a lista dos provérbios mais conhecidos e mais frequentemente utilizados pela generalidade dos falantes da comunidade linguística do português europeu. Para o estabelecimento do ‘mínimo paremiológico’ do português europeu foram utilizados diferentes procedimentos metodológicos ao longo de cinco anos. Numa primeira fase, coligiu-se uma base de dados digital com mais de 114.000 entradas (provérbios e variantes) e fez-se uma primeira seleção manual dos provérbios considerados mais usuais, tendo-se depois calculado a concordância entre anotadores, que foi bastante elevada. Procedeu-se também ao cálculo da frequência de provérbios e variantes (n. de ocorrências) em várias fontes: (1) em dicionários e coletâneas de provérbios; (2) num *corpus* de textos jornalísticos (*CETEMPúblico*, Santos & Rocha, 2001); (3) em manuais escolares de Português e de Português Língua não Materna; e (4) em dois motores de busca (Google e Bing). Procedeu-se ainda à aplicação de dois questionários distintos, *online*. Obteve-se assim uma lista de 318 provérbios – o mínimo paremiológico do português europeu. A lista dos 318 provérbios poderá ter diversas aplicações, quer para o desenvolvimento de instrumentos de diagnóstico ou terapia de certas patologias da linguagem, quer para a aprendizagem de português como língua estrangeira.

Palavras-chave: disponibilidade lexical; linguística de *corpus*; provérbio; mínimo paremiológico; português europeu; variação.

Establishing the paremiological minimum of European Portuguese

ABSTRACT. The main objective of this study is to present the paremiological minimum of European Portuguese, i.e., the list of the most widely known and frequently used proverbs employed by most speakers of the European Portuguese language community. To establish the paremiological minimum of European Portuguese, different methodological procedures have been applied, over a period of five years. In the first phase, a digital database with more than 114,000 entries (proverbs and variants) was collected and a first manual selection of the most common proverbs was made, and then the agreement between annotators was calculated, which was quite high. The frequency of proverbs and variants (number of occurrences) was also calculated in several sources: (1) in dictionaries and collections of proverbs; (2) in a corpus of journalistic texts (*CETEMPúblico*, Santos & Rocha, 2001); (3) in Portuguese and Portuguese as a Foreign Language textbooks; and (4) in two search engines (Google and Bing). Two distinct online questionnaires were also applied. A list of 318 proverbs – the paremiological minimum of European Portuguese – was thus obtained. The list of the 318 proverbs may have several applications, either for the development of diagnostic or therapeutic tools for certain language pathologies, or for learning Portuguese as a foreign language.

Keywords: lexical availability; *corpus* linguistics; proverbs; paremiological minimum; European Portuguese; variation.

Received on February 7, 2020.

Accepted on June 4, 2020.

Introdução

O interesse em determinar um conjunto de provérbios comum, conhecido pela comunidade linguística que os utiliza no dia a dia, tem sido partilhado entre profissionais de diferentes áreas do conhecimento.

Mieder (1992, 1994) reflete sobre investigações desenvolvidas por psicólogos e outros estudiosos, no âmbito de determinar o conhecimento que as pessoas têm de provérbios, e explica quais os provérbios

particularmente bem conhecidos nos Estados Unidos. O autor acrescenta que estes provérbios fazem parte da cultura dos falantes do inglês, e que os mais comuns destes formam um mínimo paremiológico, usando as próprias palavras do autor, *'minimum of proverbial knowledge'*, que é preciso conhecer de forma a se poder comunicar efetivamente em inglês. Acrescenta ainda que este assunto é de grande importância para os lexicógrafos que estão envolvidos na escrita de dicionários de língua estrangeira, ou mesmo para os professores que ensinam inglês. Para o autor, a literacia e a cultura paremiológica andam de mãos dadas.

Permjakov (1973) foi, que saibamos, um pioneiro no sentido de ser o primeiro a estabelecer o conjunto de provérbios mais conhecidos e frequentemente utilizados de uma comunidade cultural, mais especificamente da população de Moscovo. O autor determinou um mínimo paremiológico russo, que incluía cerca de 300 provérbios.

Outros estudos têm vindo a ser desenvolvidos neste âmbito, em várias línguas, designadamente a série 'Mínimo paremiológico', publicada na *Biblioteca paremiológica y fraseológica* (Ruiz-Ayúcar & Muñoz, 2016), do Centro Virtual Cervantes. Esta série inclui um conjunto de estudos sobre 'paremias'. O seu objetivo principal é observar o uso dessas 'paremias' do espanhol (de Espanha) na atualidade e determinar quais destas é que devem estar presentes no processo de ensino-aprendizagem do espanhol. Um outro objetivo do projeto é estabelecer uma 'correspondencia' entre esse mínimo paremiológico do espanhol e outras línguas de trabalho, nomeadamente o francês, o alemão, o italiano, o português, entre outras. Com este termo 'correspondencia' pretende-se designar a relação entre fraseologismos de diferentes línguas, isto é, a equivalência entre fraseologismos de uma língua de partida e uma língua de chegada, de modo a evitar a possível perda de informação, muitas vezes associada às traduções literais das expressões deste género (Ruiz-Ayúcar & Muñoz, 2016). No quadro desse projeto, tem vindo a desenvolver-se um Mínimo Paremiológico do Português (ainda inédito), tendo como objetivo geral determinar o número de 'paremias' que fazem parte da competência linguística dos falantes nativos dos países que têm o português como língua oficial. Esse Mínimo do Português Europeu (Portugal continental e insular) será composto por 329 'paremias' e será publicado¹ na série 'Mínimo paremiológico'.

Foi com base na metodologia utilizada para determinar o mínimo paremiológico do espanhol que foram desenvolvidos os estudos análogos para a determinação dos 'minima' paremiológicos das restantes línguas que fazem parte do mesmo projeto. Os autores utilizaram diferentes estratégias para poder estabelecer um repertório paremiológico mínimo, por exemplo, solicitando aos informantes que lhes indicassem espontaneamente quais as 'paremias' populares que conhecem; ou apresentando-lhes uma lista da qual estes deveriam indicar quais as 'paremias' que conhecem e se as usam; ou ainda apresentando-lhes um conjunto de 'paremias' truncadas, mostrando apenas o seu primeiro membro, com o objetivo de os falantes as completarem. Após delimitarem o mínimo paremiológico, os autores verificaram em diferentes fontes documentais (imprensa, literatura e ensaios) e em contextos orais (meios de comunicação e informantes) se essas 'paremias' eram efetivamente utilizadas pelos falantes.

Gostaríamos, no entanto, de ter visto descritos os procedimentos para a determinação dos 'minima' paremiológicos das demais línguas que fazem parte daquele projeto (especificamente para o caso do português), com os quais já foram estabelecidas 'correspondencias' com as 'paremias' do espanhol.

No *Refranero multilingue*² encontram-se 1.601 provérbios portugueses. Na ficha técnica destes provérbios é indicado se cada um destes é *'De uso actual'*; *'Muy usado'*; *'Poco usado'* ou *'En desuso'*. Não há, contudo, qualquer informação sobre como foi calculada esta informação. Foi esta uma das lacunas no conhecimento paremiológico do português que o nosso estudo visa preencher.

Diferentes métodos têm vindo a ser utilizados para a determinação dos 'minima' paremiológicos de diferentes comunidades culturais. Falaremos de alguns desses trabalhos já a seguir.

Đurčo (2015) descreve as diferentes pesquisas que desenvolveu, de modo a estabelecer um 'Paremiological Optimum' para o eslovaco (Đurčo, 2004, 2005a, 2005b, 2006, 2007).

O autor, numa primeira fase, procurou determinar quais das unidades paremiológicas registadas nas recolhas lexicográficas eram familiares aos falantes de eslovaco e eram por estes usadas. Considerando o elevado número de provérbios e variantes que se encontram nestas recolhas, o autor procurou limitar o *corpus* de trabalho. Para tal, foram necessários 5 passos: (i) a análise do registo lexicográfico de provérbios em dicionários antigos e atuais; (ii) a determinação de um conjunto principal de provérbios, após a redução

¹ Informação recuperada de <http://www.ileel.ufu.br/ileel/?p=15619> 33

² Recuperado em <https://cvc.cervantes.es/lengua/refranero/>

do material paremiológico, feita por diferentes especialistas (linguistas, lexicógrafos, paremiólogos e etnólogos); (iii) a aplicação de um questionário *on-line* com o intuito de verificar se os inquiridos conheciam os provérbios selecionados pelos especialistas e se tinham familiaridade com esses provérbios; (iv) a análise das ocorrências em ‘corpora’ dos provérbios e suas variantes selecionados pelos inquiridos no questionário; e (v) a comparação e correlação dos provérbios considerados mais familiares pelos inquiridos com os provérbios mais frequentes em ‘corpora’.

O questionário desenvolvido por este autor apresentava 2.834 provérbios e deveria ser descarregado pelos usuários, devidamente preenchido, e depois enviado via *internet*. O questionário apresentava uma parte inicial destinada a recolher a informação pessoal de cada respondente, nomeadamente o sexo, a idade, a escolaridade, as regiões onde viveu e a região onde vivia. Depois de preenchidos os dados pessoais, o respondente deveria escolher uma palavra-passe e, só então, ser-lhe-iam apresentados os provérbios, um de cada vez, e 4 opções de resposta: (i) ‘Conheço e uso o provérbio’; (ii) ‘Conheço, mas não uso o provérbio’; (iii) ‘Não conheço o provérbio, mas compreendo-o’³; (iv) ‘Não conheço o provérbio e não o compreendo’. Uma quinta opção também era apresentada, na qual o respondente poderia colocar uma variante que conhecesse desse provérbio. O programa permitia ainda que fossem alteradas as respostas dadas e que fossem adicionados, no final do questionário, outros provérbios que o respondente conhecesse e usasse, e que não tivessem sido contemplados pelo questionário. Uma outra característica deste questionário era permitir que o respondente interrompesse o seu preenchimento sempre que pretendesse.

O questionário obteve 42 respostas. Os respondentes, 24 mulheres e 18 homens, eram de diferentes regiões da Eslováquia e apresentavam uma idade média de 43 anos.

Apenas um pequeno conjunto de provérbios era conhecido e usado pelos respondentes (cerca de 16%, incluindo diferentes variantes). Com base nas respostas, foram formadas 5 categorias de provérbios: (i) o primeiro conjunto representa os provérbios que faziam parte do vocabulário ativo dos respondentes, e incluía aproximadamente 250 provérbios (uso declarado por mais de 50% dos entrevistados); (ii) um outro conjunto representa os provérbios que eram familiares aos respondentes, mas que não faziam parte do seu vocabulário ativo, e englobava 50 provérbios (em conformidade com mais de 50% dos inquiridos); (iii) o maior conjunto corresponde aos provérbios que ‘não’ eram conhecidos pelos respondentes, mas que eram por estes interpretados, e apresentava cerca de 1.900 provérbios; (iv) 100 dos 2.834 foram selecionados pelos respondentes como ‘não conhecidos’ e ‘não interpretáveis’ (em conformidade com mais de 50% dos inquiridos); e, finalmente, (v) apenas 2% dos respondentes apresentou outras variantes de provérbios que conhecia. Com base nos resultados obtidos foram então selecionados os provérbios que foram integrar o mínimo paremiológico.

De seguida, o conjunto de provérbios que fazem parte do vocabulário ativo dos respondentes foi comparado com a sua frequência de ocorrência no *Slovak National Corpus*⁴. Os resultados obtidos demonstram que 4 dos provérbios selecionados pelos respondentes não apresentavam qualquer ocorrência no *corpus*; 11 dos provérbios que foram considerados familiares pelos respondentes apresentavam um baixo número de ocorrências; e apenas 17 dos provérbios selecionados apresentam mais de 50 ocorrências no *corpus* em análise. A correlação entre, por um lado, o nível de conhecimento/familiaridade com os provérbios indicados pelos inquiridos e, por outro lado, os valores de frequência dos provérbios no *corpus* permitiu ao autor estabelecer o mínimo paremiológico da língua eslovaca.

O mesmo método foi usado também para a língua eslovena (Meterc, 2012, 2014), o que permitiu comparar formal, semântica e pragmaticamente (*‘diasystemic’*) a equivalência entre provérbios eslovacos e eslovenos, selecionados com base numa mesma metodologia empírica (Đurčo & Meterc, 2014).

Em suma, o mínimo paremiológico de uma língua deverá não só conter os provérbios que são parte de um fundo comum dessa língua, como também aqueles que são efetivamente utilizados pelos falantes.

Contudo, falar de provérbios é falar de variação (Chacoto, 1994), o que implica, em contraponto, a noção de que, intersetando as múltiplas variantes de um provérbio está uma unidade conceptual e pragmática, uma ‘invariante’, a que chamamos ‘unidade paremiológica’ (UP). Ainda que a expressão ‘unidade paremiológica’ não seja completamente inusitada - veja-se por exemplo a discussão terminológica de Gasanova e Taibova (2016), que compara *‘saying’*, *‘proverb’* e o termo *‘paroemi (a paremiological unit)’* -

³ Esta opção não está contemplada no nosso estudo, tendo apenas sido apresentadas as opções de resposta ‘Conheço e uso’, ‘Conheço mas não uso’, ‘Não conheço’ (logo, não uso). Esta opção (iii) introduz uma variável que, quanto a nós, seria difícil controlar: a representação do significado do provérbio.

⁴ Recuperado em www.Korpus.sk.

acreditamos que o conceito de UP que aqui definimos está subjacente, de forma mais ou menos implícita, em numerosos trabalhos da literatura. No entanto, a sua utilização aqui implica, quanto a nós, um esforço na determinação dos critérios (formais, semânticos e pragmáticos) que permitem considerar diferentes formas como variantes de uma mesma UP (Reis & Baptista, 2016c).

No estabelecimento do mínimo paremiológico (MP) deverá ter-se em conta, pois, o fenómeno da variação, já que, de entre as variantes associadas à mesma UP, provavelmente uma é mais usual (eventualmente mais do que uma).

O MP deve ainda ser considerado um objeto dinâmico, uma lista que deverá ser atualizada ao longo do tempo, de modo a representar o conhecimento linguístico de uma dada comunidade, num dado momento.

Seguidamente, apresentaremos a base de dados digital de provérbios que serviu como ponto de partida para o estabelecimento do mínimo paremiológico do português europeu por nós estabelecido.

Para a constituição de uma base de dados digital de provérbios (Reis & Baptista, 2016a), procedeu-se à digitalização de quatro obras de cariz paremiológico (dicionários e coletâneas; Costa, 1999, Machado, 1996, Moreira, 1996, Parente, 2005). Numa primeira fase, atribuiu-se um código convencional a cada uma das expressões, de forma a que fossem identificadas univocamente consoante a fonte de onde foram recolhidas; depois, foi necessário efetuar um conjunto de transformações na listagem inicial, já que continha erros procedentes da digitalização; foi ainda necessário uniformizar essa listagem, de forma a que a pudessemos tratar com ferramentas de processamento de texto - retirando espaços em branco e sinais de pontuação espúrios, retirando as notas e informações contidas no texto original, e outras informações adicionais, e.g. GLP_JPM05358 “Cansa quem dá, não cansa quem toma (XVII, Delic, p. 189)” (e.g. Machado, 1996, p. 134, grifo nosso); LP_SP13673 “Em S. Lourenço (10/8), vai à vinha e enche o lenço” (e.g. Parente, 2005, p. 247, grifo nosso); LP_SP04377 “Ano de salmões, ano de paixões (Minho)” (e.g. Parente, 2005, p. 84, grifo nosso). Numa fase posterior, iniciámos outra tarefa, que consistiu em assinalar todas as expressões que não são provérbios, (nomes compostos, adjetivos compostos, frases fixas, ou outras). Neste momento, já assinalámos cerca de 4.600 destas expressões. Dada a dimensão da base de dados - mais de 114.000 entradas - esta tarefa ainda se encontra em curso.

Paralelamente a esta tarefa, estão também a ser desdobradas algumas entradas que apresentam variantes e.g. LP_SP04322 “Animal de bico nunca fez o ‘dono (amo)’ rico” (e.g. Parente, 2005, p. 84, grifo nosso); GLP_JPM05553 “Casa de pai, vinha de avô. Var.: Casa de pai, vinha de avó” (e.g. Machado, 1996, p. 138). Presentemente, já desdobrámos 277 entradas na nossa base de dados e temos cerca de 5.600 entradas assinaladas para desdobrarmos.

A base de dados de provérbios assim constituída foi o ponto de partida para a determinação do mínimo paremiológico do português europeu.

Determinando um mínimo paremiológico para o português europeu

Da interseção dos vários métodos, constituímos o mínimo paremiológico do português europeu (MP), que inclui 318 provérbios. Trata-se de uma lista dinâmica, passível de ser modificada/ajustada, em função de novos dados que possam surgir. Seguidamente, descreveremos, de modo sucinto, cada uma das etapas que nos permitiram determinar a lista dos provérbios muito usuais.

Após a constituição da base de dados digital de provérbios, procedeu-se à inventariação das expressões mais frequentes (Reis & Baptista, 2016c). A frequência dos provérbios na base de dados pareceu-nos ser um indicador dos provérbios mais usuais, uma vez que foi constituída a partir de diferentes recolhas e dicionários de diferentes autores.

Para podermos agrupar esses provérbios, procurámos identificar as palavras-chave comuns entre eles, de forma semiautomática e com base em ferramentas de processamento da linguagem natural. Por palavras-chave entendemos os elementos lexicais plenamente significativos. Por exemplo, o provérbio ‘Antes tarde do que nunca’ tem como palavras-chave [‘tarde nunca’]. Para esse fim, removemos todas as *stopwords* ou seja, as palavras gramaticais que, de um modo geral, não contribuem para a identificação dos provérbios (sobretudo, determinantes, pronomes, conjunções e preposições).

Verificou-se, contudo, que as variantes de um mesmo provérbio nem sempre apresentam as mesmas palavras-chave, resultado do processo automático de construção dessas mesmas chaves, como acontece em:

(1a) Antes tarde que nunca [tarde nunca] (e.g. Parente, 2005, p. 90).

(1b) Mais vale tarde do que nunca [mais vale tarde nunca] (e.g. Parente, 2005, p. 333).

(1c) Melhor é tarde que nunca [melhor é tarde nunca] (e.g. Parente, 2005, p. 345).

(1d) Vale mais tarde do que nunca [vale mais tarde nunca] (e.g. Parente, 2005, p. 730).

Há, também, casos em que os provérbios apresentam as mesmas palavras-chave, mas que estas correspondem a provérbios diferentes, o que tem de ser analisado manualmente e caso a caso. Veja-se, a propósito, os exemplos (2) e (3) cuja chave é [‘e e’]:

(2) Essa é que é essa (e.g. Parente, 2005, p. 258).

(3) O meu é meu, o teu é nosso (e.g. Parente, 2005, p. 463).

Note-se que o número de variantes que um provérbio apresenta pode não estar relacionado com a sua disponibilidade lexical. Dito de outra forma, o facto de um provérbio apresentar muitas variantes não faz dele necessariamente um provérbio usual. Há provérbios muito usuais que apresentam uma tal fixidez dos seus constituintes que não admitem qualquer variação, como por exemplo o provérbio ‘Contas são contas’. Em contrapartida, o provérbio ‘Não se apanham trutas a bragas enxutas’ tem pelo menos 12 variantes mas não é usual.

Desta forma, para podermos agrupar as diferentes variantes de provérbios numa mesma unidade paremiológica foi necessário utilizar outras metodologias, de que falaremos mais adiante neste trabalho.

Em complemento desta abordagem, efetuou-se uma seleção dos provérbios considerados mais usuais, percorrendo-se a base de dados de provérbios (Reis & Baptista, 2017a). Esta seleção foi feita por dois anotadores, os autores deste artigo, que são falantes nativos do português europeu, linguistas de profissão, e com vasto conhecimento de provérbios.

Tinha-se como hipótese subjacente a esta metodologia que os provérbios selecionados por ambos os anotadores seriam ‘muito usuais’ (‘nível 2’), os selecionados por somente um dos anotadores seriam (apenas) ‘usuais’ (‘nível 1’) e os que não fossem assinalados por nenhum dos anotadores seriam ‘não usuais’ (‘nível 0’).

Verificámos uma elevada concordância entre os provérbios anotados, uma vez que 276 provérbios foram selecionados pelos dois anotadores - ‘nível 2’ (‘Quem canta seus males espanta’); 566 provérbios foram selecionados por apenas um dos anotadores - ‘nível 1’ (‘Em tempo de guerra não se limpam armas’) e os restantes (+113.000), não foram selecionados por nenhum dos anotadores - ‘nível 0’.

Contudo, estes números não são finais, uma vez que há repetições que terão de ser assinaladas e variantes de um mesmo provérbio que deverão ser agrupadas. Ainda assim, o facto de ter havido uma elevada concordância entre os anotadores poderá ser um indicador de que os provérbios escolhidos são efetivamente muito usuais.

De modo a podermos validar esta seleção, foi desenvolvido um inquérito por questionário, do qual falaremos de seguida.

O inquérito por questionário, a que chamámos ‘Provérbios Usuais’⁵ (Reis & Baptista, 2017a), teve como primeiro objetivo verificar se um conjunto de provérbios eram ou não conhecidos e utilizados pelos falantes nativos do português europeu. Para este inquérito foi selecionada uma amostra estratificada de 100 provérbios selecionados aleatoriamente, em três estratos. Desta amostra constam 25 provérbios de ‘nível 2’; 25 de ‘nível 1’ e 50 de ‘nível 0’. Era pedido aos informantes que indicassem para cada provérbio se conheciam e usavam (‘nível 2’), se conheciam mas não usavam (‘nível 1’) ou se não conheciam (‘nível 0’).

Este questionário recebeu 738 respostas e permitiu-nos confirmar, em geral, a disponibilidade lexical que os anotadores atribuíram aos provérbios da base de dados.

Por um lado, pôde-se confirmar a seleção manual relativamente aos provérbios indicados como ‘muito usuais’ (‘nível 2’) e ‘não usuais’ (‘nível 0’); por outro lado, os provérbios que tinham sido assinalados por apenas um dos anotadores - ‘usuais’ (‘nível 1’) - terão de ser analisados mais minuciosamente, de forma a se considerar se deverão integrar a lista dos provérbios ‘muito usuais’ ou passar para a classe dos ‘não usuais’.

Com base nas respostas obtidas no questionário, calculou-se o índice de correlação de Pearson entre o nível indicado na resposta mais votada e o nível que lhe fora atribuído como referência (com base nos dois anotadores especialistas). Dito de outra forma, comparou-se a resposta mais votada pelos respondentes para cada provérbio com a classificação dos dois anotadores. Considerando os 100 provérbios do questionário, verificou-se que há uma correlação ‘bastante alta’ (0,77) entre o nível indicado na resposta mais votada e o nível que lhe fora atribuído como referência. Nesta interpretação do índice de correlação de Pearson,

⁵ Recuperado de <https://goo.gl/forms/NvfkLXMTdCb5dA0t2>

seguimos a proposta de Parsian (2015). Isto parece indicar que o nível atribuído pelos anotadores não difere substancialmente do que é atribuído pelos respondentes do inquérito, o que confirma, pelo menos parcialmente, a classificação inicial feita pelos anotadores para os restantes provérbios selecionados, sobretudo os de ‘nível 1’.

Foram ainda utilizadas outras formas de validação da disponibilidade lexical dos provérbios que apresentámos neste questionário. É desses dados que falaremos em seguida.

Procurou-se o conjunto de provérbios apresentados no questionário de que falamos atrás nos motores de busca *Google* e *Bing*. A pesquisa restringiu-se a páginas do domínio de topo.pt e escritas em português europeu. Observou-se que os resultados, mais uma vez, confirmavam, no geral, a seleção manual dos anotadores, uma vez que os provérbios anotados como ‘muito usuais’ (‘nível 2’) eram efetivamente os que ocorriam com mais frequência na *web*. Pontualmente, verificou-se que alguns provérbios deste nível ocorrem um número reduzido de vezes, o que levou a reavaliar a sua classificação inicial (ou como de ‘nível 1’ e, mais raramente, como de ‘nível 0’). Quanto aos provérbios de ‘nível 0’, considerados ‘não usuais’, confirma-se claramente que ocorrem muito poucas vezes ou não ocorrem de todo na *web*. A situação mais complexa verifica-se nos provérbios de ‘nível 1’, marcados como apenas ‘usuais’, já que alguns apresentam valores de frequência idênticos aos dos provérbios de ‘nível 2’ (‘muito usuais’) o que levou à análise mais cuidadosa das suas ocorrências e, por vezes, à sua reclassificação no ‘nível 2’. Inversamente, certos provérbios de ‘nível 1’ parecem não ocorrer na *web*, o que poderá significar ser necessário avaliar melhor, por recurso a outros métodos, a sua disponibilidade lexical.

No geral, verifica-se que há uma correlação bastante alta entre os resultados da frequência dos dois motores de busca e a classificação inicial feita manualmente pelos anotadores.

Não sendo os motores de busca uma ferramenta suficientemente fiável, por si só, recorreremos ao *corpus* de texto jornalístico *CETEMPúblico* (Santos & Rocha, 2001) para aferir a frequência destes provérbios. Foram utilizadas duas abordagens diferentes com base em ferramentas de processamento da linguagem natural, e com recurso a transdutores de estados finitos (FST) (Paumier, 2016). Uma destas abordagens consistiu na construção manual de ‘FST’ que representam cada ‘unidade paremiológica’ (Figura 1).

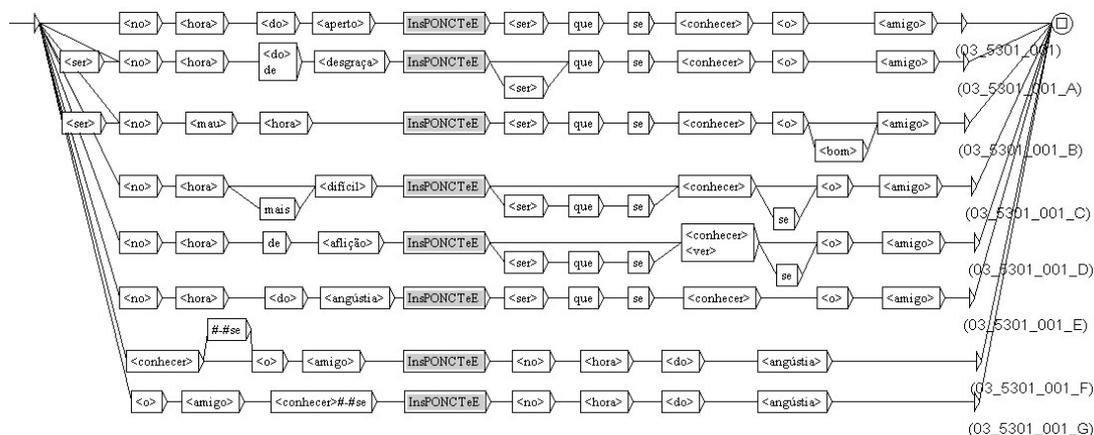


Figura 1. Transdutor do provérbio “Na hora do aperto é que se conhecem os amigos” (e.g. Parente, 2005, p. 369).

Fonte: Elaborada pelos autores.

Tentámos representar nestes FST todas as variantes de uma UP que encontrámos nas nossas fontes. Este é, no entanto, um processo em constante atualização, uma vez que poderão ser adicionadas outras variantes que não estejam atestadas na nossa base de dados. Esta abordagem tem a vantagem de poder recorrer à variação morfossintática, permitindo capturar assim todas as variantes que apresentem este tipo de variações (‘Conhece-se o amigo’ na hora da angústia; ‘Conhecem-se os amigos’ na hora da angústia). Em contrapartida, este método não permite capturar expressões que não estejam previstas no grafo. Tendo em conta que a construção destes transdutores é uma tarefa demorada e que ainda se encontra em curso, foi necessário empregar uma outra abordagem, mais expedita, que nos permitisse capturar os provérbios que ainda não tivessem sido abrangidos pela primeira. Esta consistiu no desenvolvimento de um conjunto de transdutores com base nas palavras-chave de cada provérbio (Figura 2). Este método tem como vantagem o

facto de os ‘FST’ poderem ser gerados automaticamente a partir da base de dados. Ainda que não produza resultados tão precisos como a primeira abordagem, é um método com maior abrangência (*recall*).

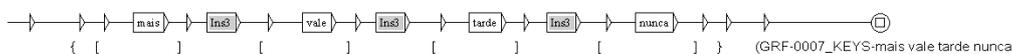


Figura 2. Transdutor do provérbio “Mais vale tarde do que nunca” (e.g. Parente, 2005, p. 333).

Fonte: Elaborada pelos autores.

O resultado da aplicação destes dois conjuntos de ‘FST’ no *corpus CETEMPúblico* (Santos & Rocha, 2001) permitiu-nos encontrar 30 provérbios diferentes, correspondendo a 13 unidades paremiológicas. Permitiu-nos ainda confirmar alguns dos provérbios mais utilizados e refletir sobre as variantes empregues. Todos os provérbios encontrados fazem parte do conjunto dos provérbios usuais (‘nível 2’), de acordo com a referência da anotação manual.

Duas outras pesquisas foram realizadas tendo, de entre os objetivos estabelecidos, o de determinar os provérbios (e as suas variantes) mais difundidos em conjuntos de manuais escolares (Reis & Baptista, 2016b, 2017b) quer de português língua materna (LM), quer de português língua não materna (PLNM), com recurso a ferramentas de processamento de linguagem natural (‘FST’ que representam as unidades paremiológicas e ‘FST’ com as palavras-chave).

Duas expectativas orientaram as pesquisas: (i) por um lado, esperava-se que os provérbios usuais aparecessem repetidos ou no mesmo manual ou em diferentes manuais; (ii) por outro lado, julgava-se que os provérbios presentes nos manuais corresponderiam, na grande maioria, a provérbios ‘usuais’ ou ‘muito usuais’, independentemente da sua frequência nestas obras.

A ideia por detrás desta segunda expectativa está relacionada com o facto de se pressupor que a escolha de provérbios realizada pelos autores destes manuais, de algum modo poderia refletir a sua disponibilidade lexical, já que é naturalmente desejável que os provérbios a apresentar neste tipo de instrumentos didáticos, dirigidos a um público de aprendentes, correspondam de forma mais direta a expressões quotidianas, a traços mais vinculados da cultura portuguesa, ou seja, a provérbios usuais na língua e na cultura portuguesas.

Para a pesquisa de provérbios em manuais de português (LM), recorreu-se a um *corpus* de 38 manuais escolares, da Porto Editora, do 5 ao 12º anos, com 2,9 milhões de palavras, que já se encontrava digitalizado (Curto, 2014), tendo-se encontrado 272 provérbios, que correspondiam a 192 unidades paremiológicas distintas. Neste caso, não foi possível estabelecer uma lista dos provérbios mais usuais, uma vez que poucos se repetiram. Nos casos em que houve repetições, verificou-se que, por vezes, um mesmo provérbio era utilizado em mais do que um exercício no mesmo manual, o que poderá ter influenciado ou enviesado os resultados obtidos.

Um outro aspeto a ter em consideração é o facto de alguns provérbios encontrados ocorrerem em textos literários que constam destes manuais ou serem referidos a propósito destes textos. Uma vez que os mesmos provérbios se poderão repetir em diferentes manuais escolares do mesmo ano de escolaridade, e dado que os textos estudados são os mesmos, essa repetição não é necessariamente significativa. Veja-se, por exemplo, o provérbio: “(4) Voz do povo, voz de Deus” (e.g. Garrido, Duarte, Rodrigues, Afonso, & Lemos, 2000, s.p.), que se encontra em quatro dos cinco manuais do 11º ano deste *corpus*, no texto dramático *Frei Luís de Sousa*, de Almeida Garrett “MARIA - Voz do povo, voz de Deus, minha senhora mãe: eles que andam tão crentes nisto, alguma coisa há-de ser (ato I, cena III)”.

Ao compararmos os resultados de frequência dos provérbios nestes manuais com os níveis de disponibilidade lexical dos provérbios definidos manualmente pelos anotadores, verificámos que 50% dos provérbios encontrados são ‘não usuais’ (de ‘nível 0’), 28% são ‘muito usuais’ (de ‘nível 2’) e 22% são apenas ‘usuais’ (de ‘nível 1’). O número de provérbios ‘não usuais’ encontrado é igual à soma dos marcados como ‘usuais’ e ‘muito usuais’. Este resultado não era o esperado, considerando que era desejável que os manuais representassem os provérbios (mais) usuais na língua e na cultura portuguesas. Ao verificarmos os índices de frequência destes provérbios ‘não usuais’ na *web*, pudemos confirmar que são efetivamente provérbios ‘muito pouco usuais’, uma vez que 26% não apresentaram quaisquer ocorrências e 66% destes provérbios tiveram menos de 30 ocorrências. Recorde-se que o limiar de frequência considerado para os provérbios de ‘nível 2’ na *web* era 50). Ainda assim, 8% dos provérbios encontrados apresentaram mais de 30 ocorrências, pelo que foram eventualmente reclassificados nos outros dois níveis de disponibilidade lexical (‘nível 1’, ‘nível 2’).

Um estudo semelhante foi desenvolvido num conjunto de manuais de português língua não materna (PLNM), publicados pela LIDEL, de vários níveis de proficiência, com aproximadamente de 250 mil palavras. Este conjunto de textos foi especificamente digitalizado para esta fase do estudo.

Os resultados desta pesquisa permitiram identificar 34 provérbios, correspondendo a 20 unidades paremiológicas distintas. À semelhança do que verificámos nos manuais de português (LM), poucos provérbios se repetiram (apenas 7). Desta forma, não foi possível determinar se alguns provérbios seriam mais usuais nestes manuais. Pudemos, no entanto, observar que todos os provérbios encontrados podiam ser considerados ‘usuais’ ou ‘muito usuais’, com base nos dados de disponibilidade lexical, isto é, considerando a anotação manual que nos serviu de referência.

Outro questionário que desenvolvemos, o ‘Questionário de avaliação da qualidade de jogos proverbiais’⁶ (Reis, Pompili, Abad, & Baptista, 2017), consistiu na aplicação de 100 jogos – *cloze questions* - com provérbios (estímulos), que apresentavam quatro opções de resposta (uma resposta certa e três ‘distratores’), todas linguisticamente motivadas. Destes jogos, 12 apresentavam duas respostas corretas. O questionário era constituído por duas partes distintas. Na ‘Parte 1’, apresentava-se o conjunto de jogos com provérbios que se pedia que fossem resolvidos; na ‘Parte 2’, era solicitada a opinião dos respondentes relativamente à parte anterior.

O primeiro objetivo deste questionário consistia em avaliar a qualidade de um conjunto de jogos com provérbios, com vista a vir a inseri-los numa plataforma virtual de diagnóstico e terapia de algumas patologias da linguagem. No entanto, também nos permitiu validar, de modo indireto, a seleção dos provérbios usuais feita pelos dois anotadores e já descrita anteriormente, uma vez que a grande parte dos provérbios que serviram neste questionário (selecionados aleatoriamente e todos de ‘nível 2’) não integrara a amostra do questionário ‘Provérbios Usuais’ (apenas 21 se repetiram). Ao utilizarmos um outro conjunto de provérbios já assinalado como ‘muito usuais’, pudemos validar a classificação dos anotadores sempre que os jogos de provérbios foram corretamente resolvidos por um elevado número de respondentes.

Responderam a este questionário 164 pessoas e, em média, a taxa de acerto foi de 93%. Estes dados confirmam a apropriada classificação dos provérbios apresentados como ‘muito usuais’, pois em geral os falantes conheciam e sabiam usar os provérbios. Nos casos em que apresentámos mais do que uma solução correta, a maioria dos respondentes selecionou a opção que consideramos ser a variante mais usual dessa UP.

A ‘Parte 2’ do questionário permitiu-nos ainda validar a metodologia seguida na seleção dos ‘distratores’, tendo em vista uma eficiente construção dos exercícios. Em concreto, perguntava-se se a solução dos exercícios se encontrava entre as opções indicadas e se o informante conseguia completar facilmente o provérbio com base nas opções de resposta apresentadas, mesmo quando não o conhecia.

Na primeira questão, a alta percentagem de pessoas que indicou que a solução se encontrava ‘sempre’ ou ‘a maioria das vezes’ (53,7 e 44,5%, respetivamente) nas opções de resposta apresentadas, permitiu concluir, em geral, que os provérbios tinham sido reconhecidos pelos informantes. A segunda questão, quanto à facilidade em completar os provérbios com base nas opções de resposta apresentadas está ligada ao facto de essas respostas terem todas um racional subjacente, que permite determinar por analogia a resposta certa. Verificou-se que 20,7% dos respondentes selecionou a opção ‘sempre’ e 68,9% a opção ‘a maioria das vezes’, o que confirma a adequação dos ‘distratores’ produzidos.

Constituição da lista de provérbios usuais – Mínimo paremiológico

Os resultados obtidos em cada uma das etapas, já descritas anteriormente, permitiram-nos estabelecer a lista final de provérbios muito usuais - o mínimo paremiológico do português europeu -, que apresentamos em Anexo. De seguida, descreveremos com algum pormenor este processo de seleção.

Primeiramente, retomou-se a lista de provérbios que foi objeto de seleção manual pelos anotadores a partir da base de dados de provérbios. Foi necessário verificar todos os provérbios ou variantes repetidas que tinham sido selecionados inadvertidamente; eliminar as repetições; e escolher de entre as variantes repetidas a mais usual. Em alguns casos considerou-se apenas um dos membros dos provérbios, cuja frequência é muito superior à utilização da forma mais longa com os dois membros, como sucede, por exemplo, em “Não há regra sem exceção [nem mulher sem senão]” (e.g. Parente, 2005, p. 392). Parte deste trabalho já tinha sido feita, mas agora esta análise incidiu não só nos provérbios classificados como ‘muito usuais’ (de ‘nível 2’), como também nos considerados ‘usuais’ (de ‘nível 1’).

⁶ Recuperado de <https://goo.gl/forms/H0NxEKTxUBxFFZzB2>

Atentemos primeiro à lista de provérbios de ‘nível 2’. Nesta fase, e sob um olhar mais crítico, resolvemos excluir desta lista 44 provérbios que tinham sido assinalados inicialmente pelos dois anotadores como muito frequentes. Esta decisão assentou em diferentes critérios: na baixa frequência de uso destes provérbios, que foi verificada na *web*, em *corpora* de texto jornalístico e de manuais escolares; e nos resultados obtidos pela aplicação de questionários, dos quais voltaremos a falar, de seguida.

A aplicação dos questionários permitiu-nos excluir, com segurança, alguns provérbios, como por exemplo: “(1) A preguiça morreu de sede à beira da água” (e.g. Parente, 2005, p. 49). Este provérbio, que consta do questionário ‘Provérbios Usuais’⁷ (Reis & Baptista, 2017a), não foi reconhecido por 45,1% dos inquiridos, que selecionaram a opção ‘Não conheço’.

Em alguns casos, a variante apresentada poderá ter tido alguma influência sobre as respostas fornecidas, apesar da advertência feita no início do questionário para que fossem consideradas outras variantes além da que se mostrava: ‘interessa-nos apenas que nos indique se conhece e/ou usa os provérbios (ou uma das suas variantes) que lhe apresentamos’. Por essa razão, nos casos em que os resultados do questionário nos pareceram menos claros, complementámos esta informação de natureza introspectiva com dados de frequência na rede.

Nesta fase, a frequência de provérbios foi obtida com base no motor de busca *Google*, uma vez que já se tinha verificado que os resultados dos motores de busca *Google* e *Bing* eram idênticos/comparáveis. Uma vez mais, a pesquisa foi feita exclusivamente em páginas do domínio de topo.pt, e tendo em conta apenas as páginas escritas em português europeu (dados de janeiro de 2019). De modo a obtermos dados o mais fidedignos possível, percorremos até à última página os resultados de cada uma das pesquisas efetuadas pelo *browser*, o que nos possibilitou identificar os resultados mais relevantes, omitindo as entradas muito semelhantes às já apresentadas. Nestas pesquisas, tivemos em consideração quer a expressão *ipsis verbis* (digitando cada expressão entre aspas), quer as diferentes variantes desses mesmos provérbios (digitando cada uma das variantes manualmente ou substituindo certos elementos por *wildcard*), o que nos permitiu descartar a hipótese de o baixo uso de um dado provérbio poder dever-se à escolha da variante.

Por exemplo, considerando o provérbio “A preguiça morreu de sede à beira da água” (e.g. Parente, 2005, p. 49), somente encontramos oito ocorrências do mesmo, *ipsis verbis*. Todas as variantes (conhecidas) deste provérbio têm uma fraca expressão na *web*, quase todas na variante acompanhada de ‘andando a nadar’, com a preposição ‘a’ (‘morreu à sede’) (#5 ocorrências), e uma com a preposição ‘de’ (‘morreu de sede’). Encontrou-se ainda duas ocorrências de “A preguiça morreu de sede” (e.g. Machado, 1996, p. 50), sem um segundo hemistíquio. Não se encontrou quaisquer ocorrências de outras variantes já registadas.

Um caso particularmente expressivo é o provérbio: “(2) Não se apanham trutas a bragas enxutas” (e.g. Parente, 2005, p. 400) para o qual 91,5% dos respondentes indica ‘Não conheço’. Este foi o único caso em que os informantes indicaram que não conheciam um provérbio, inicialmente marcado como sendo de ‘nível 2’. Efetivamente, a frequência na rede do provérbio *ipsis verbis* é muito baixa (#3 ocorrências) e as variantes encontradas ocorrem predominantemente em listas de provérbios, enquanto outras variantes não ocorrem de todo.

Já o provérbio: “(3) Deus dá o frio conforme a roupa” (e.g. Machado, 1996, p. 190), apesar de ter sido marcado como ‘muito usual’, não era conhecido pela maior parte dos respondentes do questionário dos jogos, pois a maioria (68,9%) não completou corretamente o provérbio. Os ‘distratores’ apresentados ou exploravam o paralelismo formal e a rima: [deus dá o frio] ‘e o diabo dá o brio’ (45,7%); ou a rima: ‘conforme o arrepio’ (15,9%); ou a sinonímia: ‘conforme a vestimenta’ (7,3%). Apenas 31,1% acertou na opção correta.

Note-se que estes três provérbios aparecem em todas as coletâneas que utilizámos, apresentando inclusive, alguma variação. Face a estes resultados, transferimo-los para os provérbios de ‘nível 1’.

Dos 44 provérbios retirados da listagem inicial de 276 de ‘nível 2’ (seleção manual), confirmámos que 35 apresentavam índices de frequência bastante baixos, mesmo quando se considerava o conjunto de variantes, por vezes numeroso, associadas à mesma ‘unidade paremiológica’. Assim, por exemplo, para o provérbio: “(4) Cada um dança conforme a música” (e.g. Parente, 2005, p. 138) encontramos 19 ocorrências na *web*, considerando todas as 9 variantes associadas a esta unidade paremiológica (a variante ilustrada em (4) ocorre 3 vezes), mas apenas uma variante parece mais usual: “Conforme se toca, assim se dança” (e.g. Parente, 2005, p. 172) (#13 ocorrências).

Por outro lado, para o provérbio: “(5) Antes a criança chore que a mãe suspire” (e.g. Parente, 2005, p. 85), que apresenta 2 variantes, apenas esta forma ocorre na rede e com uma frequência baixa (#6 ocorrências).

⁷ Disponível em <https://goo.gl/forms/NvfkLXMTdCb5dA0t2>

Como se pode verificar, o número de variantes de uma UP não está diretamente relacionado com a sua frequência em textos.

Removemos também da lista inicial dos provérbios de ‘nível 2’ cinco variantes das mesmas unidades paremiológicas, que se encontravam repetidas, veja-se, por exemplo:

(6) A descer todos os santos ajudam (e.g. Parente, 2005, p. 21).

(6a) Para baixo todos os santos ajudam (e.g. Parente, 2005, p. 512).

Retirámos também desta lista as expressões fixas:

(7) Os dados estão lançados (e.g. Parente, 2005, p. 496).

(8) Fia-te na virgema e não corras e verás o trambolhão que levas (e.g. Parente, 2005, p. 274).

pois considerámos tratar-se de expressões, atualizáveis no discurso, e não propriamente provérbios. Removemos ainda: “(9) O bacalhau quer alho” (e.g. Parente, 2005, p. 439) por nos parecer sobretudo estar associado a uma canção popular e por não termos encontrado qualquer evidência do seu estatuto enquanto provérbio: todas as ocorrências que encontramos ou estavam relacionadas com a canção popular ou só apareciam em listas de provérbios e não usados de forma integrada no discurso. Finalmente, retirámos a expressão: “(10) O rei vai nu” (e.g. Parente, 2005, p. 480) que, apesar da sua frequência bastante elevada, nos causou dúvidas relativamente à sua classificação como provérbio. A grande maioria das ocorrências por nós encontradas de “O rei vai nu” (e.g. Parente, 2005, p. 480). correspondem a referências explícitas ao conto de Hans Christian Andersen (2006), de onde esta expressão deriva. No entanto, encontramos alguns exemplos de emprego da expressão sem qualquer relação explícita ao conto, num uso que se parece bastante com o de muitos provérbios: “É que apesar [de os] donos da cultura em Portugal se estarem a divertir bué, ‘o rei vai ‘mesmo’ nu’.⁸

Tal como foi necessário ajustar a lista de provérbios de ‘nível 2’, também foi indispensável efetuar alterações na lista de provérbios de ‘nível 1’. Assim, 76 provérbios (dos 566) que tinham sido selecionados por apenas um dos anotadores foram reclassificados, tendo agora sido integrados no mínimo paremiológico, como, por exemplo, “Da discussão nasce a luz” (e.g. Parente, 2005, p. 182).

Algumas das expressões incluídas no MP poderão surpreender o leitor por serem *slogans* e/ou citações, alguns até com autor conhecido, como, “Há mar e mar, há ir e voltar” (*slogan* cunhado por Alexandre O’Neill para a campanha de sensibilização do Instituto de Socorros a Náufragos, para prevenir os afogamentos nas praias portuguesas) ou “Tudo está bem quando acaba bem” (tradução do título da peça *All’s well that ends well* de Shakespeare - (Shakespeare, 2009, s.p). Consideramos que estas expressões já são parte do fundo comum de conhecimento da língua, e que, em relação a elas, a noção de autoria já terá praticamente desaparecido, daí que sejam muitas vezes utilizadas proverbialmente. Vejamos, a propósito, os seguintes exemplos retirados da *web*:

O ditado popular que afirma que ‘[...] há mar e mar, há ir e voltar [...]’, é volta e meia ‘violado’ pelos pescadores minhotos, que acabam por morrer na faina, como hoje aconteceu em Vila Praia de Âncora, Caminha⁹.

Chegamos ao fim de mais um ano com altos e baixos, mais positivo para uns do que outros, mas ‘como diz o ditado’, ‘tudo está bem quando acaba bem’, e por isso é importante acabar 2016 em festa, como é costume, celebrando a passagem de ano¹⁰. (grifo nosso).

Os exemplos aqui apresentados estão associados a algumas das fórmulas que muitas vezes se usam para introduzir um provérbio no discurso: ‘como diz o ditado’. O facto de, por um lado, estas expressões, de que até conhecemos a autoria, serem consideradas e usadas como provérbios pelos falantes; e, por outro lado, a sua elevada frequência de uso, tal como se reflete no número de ocorrências que apresentam na *web*; levaram-nos a incluí-las no MP. Note-se, todavia, que, até por eventual desconhecimento nosso, poderá haver nesta mesma lista outras expressões que são, de facto, citações mas que não foram por nós identificadas como tal.

Ainda em relação a este conjunto de 76 provérbios, verificámos que alguns deles também ocorriam nos manuais tanto de língua materna, como de língua não materna, por exemplo, “Quem conta um conto acrescenta um ponto” (e.g. Ferreira, & Bayan, 2016, s.p.). (#7 ocorrências) e “Uma imagem vale mais que mil

⁸ <https://www.jornaltornado.pt/quando-o-rei-vai-nu-e-preciso-tomar-a-palavra/>

⁹ <https://tvi24.iol.pt/naufrajio/vila-praia-de-ancora/ha-mar-e-mar-ha-ir-e-nao-voltar>

¹⁰ <http://lolesmtg.pt/tag/passagem-de-ano/>

palavras” (Pinto, Baptista & Fonseca, 2004, s.p.). (#4 ocorrências). Tal pode ser considerado como uma confirmação da seleção no caso dos manuais de língua não materna. Já nos manuais de língua materna, esta evidência não é tão significativa.

Verificámos que a maioria dos provérbios de ‘nível 1’ que foram reclassificados para ‘nível 2’ surge em todas as coletâneas da nossa base de dados. Há, no entanto, algumas exceções que gostaríamos de referir. O provérbio “Nada dura para sempre” (Parente, 2005, p. 371). ocorre apenas em uma das coletâneas. Ao verificámos a sua frequência na *web*, constatámos que ocorre 98 vezes (sem resultados repetidos), o que consideramos uma alta frequência de uso quando comparada com a frequência de muitos outros provérbios usuais. Confirmámos ainda que esta expressão é reiteradamente usada de modo proverbial e identificada explicitamente como um provérbio, como sucede em:

‘Nada dura para sempre...’ Foi esta ‘a máxima que repeti’ muitas vezes, foi esta ‘a máxima que passei a outras pessoas’. ‘Nada dura para sempre’, o amor magoa, a vida é lixada e tantas outras coisas negras e cheias de sofrimento (grifo nosso)¹¹.

Os dados obtidos através da aplicação dos questionários também contribuíram para esta reclassificação, na medida em que muitos destes provérbios foram reconhecidos por um elevado número de respondentes. No caso do provérbio “Na hora do aperto é que se conhecem os amigos” (e.g. Parente, 2005, p. 369), que surge no questionário ‘Provérbios Usuais’¹² (Reis & Baptista, 2017a), 94,6% dos respondentes conhece o provérbio. Este provérbio apresenta diversas variantes na base de dados, surgindo em todas as coletâneas deste estudo. Daí a sua reclassificação no ‘nível 2’.

O provérbio “Quem tem unhas é que toca guitarra” (e.g. Machado, 1996, p. 555). também foi reclassificado tendo em conta os resultados dos questionários (este é um dos casos de provérbios que se repetiram em ambos os questionários). No questionário ‘Provérbios Usuais’ (Reis & Baptista, 2017a) 78,4% dos inquiridos conhece este provérbio. No caso do questionário dos jogos, 81,1% selecionou a opção correta.

Após todas as reclassificações efetuadas e ao refletirmos sobre a lista de provérbios assim estabelecida, verificámos que havia quatro outros provérbios com uma frequência bastante elevada e que deveriam ter sido incluídos no Mínimo Paremiológico do Português Europeu. São estes os provérbios: “Ano novo, vida nova”; “Contra factos não há argumentos”; “O que é bom acaba depressa” e “O ataque é a melhor defesa”, cujo número de ocorrências na *web* é 175, 91, 107 e 72, respetivamente. Constatámos que, talvez por lapso, estes não haviam sido selecionados por nenhum dos anotadores.

Considerações finais

Em conclusão, ainda que qualquer seleção tenha sempre uma componente de arbitrariedade, os cuidados metodológicos de que nos rodeámos, bem como a justificação explícita da forma como eles foram aplicados, levam-nos a considerar que este estudo produziu uma lista de provérbios muito conhecidos e muito usuais que pode ser utilizada como o mínimo paremiológico do português europeu (MP). O estudo introduz também o conceito de unidade paremiológica (UP), formalizando explicitamente a relação entre provérbios e suas variantes e delimitando o conceito de variação.

Este instrumento poderá ter diversas aplicações, por exemplo para o desenvolvimento de instrumentos de diagnóstico ou terapia de certas patologias da linguagem; ou para a aprendizagem de português, tanto como língua materna como língua estrangeira.

Recentemente em (Reis, Baptista, & Mamede, 2020), repetiu-se a experiência relativa ao cálculo da frequência de provérbios e respetivas variantes no *corpus* jornalístico *CETEMPúblico* (Santos & Rocha, 2001), usando-se a cadeia de processamento da linguagem natural (PLN) *STRING - Statistical and Rule-Based Natural Language Processing Chain* (Mamede, Baptista, Diniz, & Cabarrão, 2012)¹³. Foi assim possível validar a seleção de parte dos 318 provérbios que constam do mínimo paremiológico aqui estabelecido. Trata-se de um tipo de texto que, como refere Figueiredo (2012), nos dias de hoje não é particularmente propenso a apresentar provérbios. Ainda assim, encontrámos 457 padrões diferentes que correspondem a 909 ocorrências de provérbios. Do conjunto das unidades paremiológicas registadas no MP, 115 não foram encontradas neste *corpus*. Confirmou-se ainda a seleção das variantes de provérbios selecionadas para

¹¹ Recuperado de <https://wrestling.pt/beyond-the-mat-27-last-forever/>

¹² Recuperado de <https://goo.gl/forms/NvfkLXMTdCb5dA0t2>

¹³ Recuperado de <http://string.i2f.inesc-id.pt>

identificar as UP no MP, uma vez que as variantes que aparecem neste *corpus* correspondem também às variantes mais usuais.

Utilizamos também as mesmas gramáticas locais da STRING para processar as bases de dados de provérbios Mosaico (Cercifaf, 2002)¹⁴ e do Projeto NATURA (1994)¹⁵, disponíveis na rede. Realce-se que, intencionalmente, estas coletâneas não foram usadas para constituir a nossa base de dados de provérbios para que pudessem agora servir de *corpus* de teste. Os resultados confirmam que os provérbios selecionados para o MP são realmente expressões usuais porque, em grande parte, surgem em ambas as coletâneas. Na primeira base de dados (Mosaico), que tem 5.363 entradas, foram encontrados 358 provérbios, correspondendo a 269 unidades paremiológicas diferentes. Do MP, 49 provérbios não foram encontrados nesta base de dados ('A carne é fraca'). Na segunda base de dados (NATURA), que continha originalmente 2.293 entradas correspondentes a provérbios, foram desdobradas as entradas que tinham variantes e retiradas as repetições, resultando em 1.902 entradas diferentes. A STRING permitiu-nos encontrar aqui 407 provérbios, que correspondem a 236 UP diferentes. Há, no entanto, 82 provérbios no MP que não constam da lista de provérbios do Projeto NATURA (1994) ("A curiosidade matou o gato"). Das unidades paremiológicas encontradas nestas duas bases de dados, 221 provérbios são comuns a ambas. Os resultados obtidos parecem, pois, validar a seleção manual de provérbios muito usuais.

Num trabalho futuro, pretendemos completar as gramáticas locais da STRING, acrescentando às expressões regulares já construídas as variantes que eventualmente ainda não tenham sido representadas. Pretendemos ainda continuar a atualizar este mínimo paremiológico, com recurso a blocos interativos e a miniquestionários a serem aplicados na *web*.

Agradecimentos

Esta investigação foi parcialmente suportada por fundos nacionais através da Fundação para a Ciência e a Tecnologia (ref. UIDB/50021/2020).

Referências

- Andersen, H. C. (2006). *The complete Hans Christian Andersen fairy tales*. New York: Gramercy Books.
- Cercifaf. (2002). Mosaico. *Base de dados de provérbios*. Recuperado de <http://www.cercifaf.org.pt/mosaico.edu/1c/proverb1.txt>
- Chacoto, L. (1994). *Estudo e formalização das propriedades léxico-sintáticas das expressões fixas proverbiais* (Dissertação de Mestrado). Faculdade de Letras da Universidade de Lisboa, Lisboa.
- Costa, J. (1999). *O livro dos provérbios portugueses*. Lisboa, PT: Editorial Presença.
- Curto, P. (2014). *Classificador de textos para o ensino de português como segunda língua* (Dissertação de Mestrado). Instituto Superior Técnico, Universidade de Lisboa, Lisboa.
- Đurčo, P. (2004). Slovak proverbial minimum: the empirical evidence. In C. Földes (Ed.), *Res humanae proverbiorum et sententiarum. Ad honorem Wolfgangi Mieder* (p. 59-69). Tübingen, DE: Narr.
- Đurčo, P. (2005a). Empirisch- und Korpusbasierte Untersuchungen der Sprichwörter. *Zeitschrift für germanistische Sprach- und Literaturwissenschaft in der Slowakei*, 1(3), 47-57.
- Đurčo, P. (2005b). Paremiologické minimum slovenčiny. Výsledky a porovnanie. In R. Blatná, & V. Petkevič (Eds.), *Jazyky a jazykov, eda. Sborník k 65. narodeninám prof. PhDr. Františka Čermáka, DrSc* (p. 45-61). Praha, CZ: Ústav Českého národního korpusu.
- Đurčo, P. (2006). Methoden der sprichwortanalysen oder auf dem weg zum sprichwörter-optimum. In A. H. Buhofer, & H. Burger (Eds.), *Phraseology in motion I. Methoden und kritik* (Phraseologie und Parömiologie, n. 19, p. 3-20). Baltmannsweiler, DE: Schneider Hohengehren.
- Đurčo, P. (2007). Paremiologické optimum slovenčiny. In W. Chlebda (Ed.), *Frazeologia a jazykove obrazy świata przełomu wieków* (p. 171-177). Opole, DE: Uniwersytet Opolski.
- Đurčo, P. (2015). Empirical research and paremiological minimum. In H. Hrisztova-Gotthardt, & M. Aleksa Varga (Eds.), *Introduction to paremiology: a comprehensive guide to proverb studies* (p. 183-205). Berlin, DE: De Gruyter.

¹⁴ Recuperado de <http://www.cercifaf.org.pt/mosaico.edu/1c/proverb1.txt>

¹⁵ Recuperado de <https://natura.di.uminho.pt/~jij/pln/proverbio.dic>

- Đurčo, P., & Meterc, M. (2014). Empirične paremiološke raziskave tipov ekvivalentnosti in suprasemantičnih razlik v slovenščini in slovaščini. *Slavia centralis*, 6(2), 20-36.
- Ferreira, A., & Bayan, H. (2016). *Na Onda do Português 3*. Lisboa, PT: Lidel.
- Figueiredo, G. (2012). *O gênero proverbial na imprensa: usos e funções retóricas* (Tese de Doutorado). Universidade Federal de Pernambuco, Recife.
- Garrido, A., Duarte, C., Rodrigues, F., Afonso, F., & Lemos, L. (2000). *Práticas 11*. Lisboa, PT: Fundação Calouste Gulbenkian.
- Gasanova, M. A., & Taibova, L. Y. (2016). Revisiting paremiological units of the Tabasaran language. *Oriental Studies*, 25(3), 99-105. Doi: 10.22162/2075-7794-2016-25-3-99-105
- Machado, J. P. (1996). *O grande livro dos provérbios*. Lisboa, PT: Editorial Notícias.
- Mamede, N., Baptista, J., Diniz, C., & Cabarrão, V. (2012). STRING: an hybrid statistical and rule-based natural language processing chain for Portuguese. In *Processing of the 10th International Conference on Computational Processing of the Portuguese Language*. Coimbra, PT. Recuperado de <https://www.inesc-id.pt/ficheiros/publicacoes/8578.pdf>
- Meterc, M. (2012). Online questionnaire providing information on most well-known and well-understood proverbs in slovene language. In *Paper presented at the EUROPHRAS Conference*. Maringor, SI. Doi: 10.13140/RG.2.1.1491.7203
- Meterc, M. (2014). *Primerjava paremiologije v slovenskem in slovaškem jeziku na osnovi paremiološkega optimuma* (doktorska disertacija). Filozofska fakulteta Univerze v Ljubljani. Ljubljana.
- Mieder, W. (1992). Paremiological minimum and cultural literacy. In S. J. Bronner (Ed.). *Creativity and tradition in folklore: new directions* (p. 185-203). Logan, UT: Utah State University Press.
- Mieder, W. (1994). *Wise words (RLE folklore)*. London, UK: Routledge.
- Moreira, A. (1996). *Provérbios Portugueses*. Lisboa, PT: Editorial Notícias.
- Parente, S. (2005). *O Livro dos Provérbios*. Lisboa, PT: Editora Áncora.
- Parsian, M. (2015). *Data algorithms: recipes for scaling up with hadoop and spark*. Sebastocol, CA: O'Reilly Media.
- Paumier, S. (2016). *Unitex 3.1 - user manual*. Université de Paris- Est, Marne-la-Vallée, FR: Institut Gaspard Monge.
- Permjakov, G. L. (1973). On the paremiological level and paremiological minimum of language. *Proverbium*, 1(22), 862-863.
- Pinto, E., Baptista, V., & Fonseca, P. (2004). *Plural 11*. Lisboa, PT: Porto Editora.
- Reis, S., & Baptista, J. (2016a). 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* (p. 427-446). Universidade do Algarve, Faro.
- Reis, S., & Baptista, J. (2016b). O uso de provérbios no ensino de português. In *Processing of the 10th Interdisciplinary Colloquium on Proverbs* (p. 521-538). Tavira, PT.
- Reis, S., & Baptista, J. (2016c). Portuguese proverbs: types and variants. In G. C. Pastor (Ed.), *Computerised and corpus-based approaches to phraseology: monolingual and multilingual perspectives* (p. 208-217). Geneva, CH: Editions Tradulex.
- Reis, S., & Baptista, J. (2017a). Estimating lexical availability of European Portuguese proverbs. In R. Mitkov (Ed.), *Computational and corpus-based phraseology. EUROPHRAS 2017. Lecture Notes in Computer Science, 10596* (p. 232-244). Londres, UK: Springer International Publishing.
- Reis, S., & Baptista, J. (2017b). Os provérbios em manuais de ensino de português língua não materna. In V. Pinheiro, & G. H. Paetzold (Eds.), *STIL 2017. XI Brazilian Symposium in Information and Human Language Technology and Collocated Events. Proceedings of the Conference* (p. 247-255). Uberlândia, MG.
- Reis, S., Baptista, J., & Mamede, N. (2020). Processing Proverb Variation in (European) Portuguese: Integrating the Paremiological Minimum in STRING. In *Paper presented in 2nd Workshop on Linguistic Tools and Resources for Paraphrasing in Portuguese, collocated with International Conference on the Computational Processing of Portuguese (PROPOR 2020)*. Colégio do Espírito Santo, Évora, PT. Doi: 10.13140/RG.2.2.14935.62880

- Reis, S., Pompili, A., Abad, A., & Baptista, J. (2017). O provérbio como estímulo num terapeuta virtual. In 6^o *Simpósio Mundial de Estudos sobre o Português (Simpósio 77 - A Importância da aprendizagem lexical*. Lisboa, PT.
- Ruiz-Ayúcar, M., & Sevilla Muñoz, J. (2016). *El mínimo paremiológico: aspectos teóricos y metodológicos* (Biblioteca fraseológica y paremiológica, Série mínimo paremiológico, n. 1). Madrid, ES: Centro Virtual Cervantes, Instituto Cervantes.
- Santos, D., & Rocha, P. (2001). Evaluating CETEMPúblico, a free resource for portuguese. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics* (p. 442-449). Toulouse, FR. Recuperado de <http://www.linguateca.pt/CETEMPUBLICO/>
- Shakespeare, W., (2009). *All's Well that Ends Well: The Cambridge Dover Wilson Shakespeare* (S. Quiller-Couch & J. Dover Wilson, Eds.). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511704079

Anexo

Mínimo Paremiológico do Português Europeu

1. A beleza está nos olhos de quem a vê.	126. Ladrão que rouba ladrão tem cem anos de perdão.	240. Quem corre por gosto não cansa.
2. A carne é fraca.	127. Longe dos olhos, longe do coração.	241. Quem dá aos pobres empresta a Deus.
3. A cavalo dado não se olha ao dente.	128. Lua de mel, lua de fel.	242. Quem dá e torna a tirar, ao inferno vai parar.
4. A culpa morreu solteira.	129. Mãe há só uma.	243. Quem desdenha quer comprar.
5. A curiosidade matou o gato.	130. Mais vale cair em graça do que ser engraçado.	244. Quem diz a verdade não merece castigo.
6. A descer todos os santos ajudam.	131. Mais vale pouco que nada.	245. Quem diz o que quer, ouve o que não quer.
7. A esperança é sempre a última a morrer.	132. Mais vale prevenir do que remediar.	246. Quem é vivo sempre aparece.
8. A experiência é a mãe da ciência.	133. Mais vale tarde do que nunca.	247. Quem espera sempre alcança.
9. A falar é que a gente se entende.	134. Mais vale um pássaro na mão do que dois a voar.	248. Quem espera, desespera.
10. A fé move montanhas.	135. Manda quem pode.	249. Quem está mal que se mude.
11. A fome é o melhor tempero.	136. Mãos frias, coração quente.	250. Quem estraga velho paga novo.
12. A galinha da vizinha é sempre melhor que a minha.	137. Melhor é dobrar que quebrar.	251. Quem fala assim não é gago.
13. A idade não perdoa.	138. Muita parra, pouca uva.	252. Quem faz um cesto faz um cento.
14. A intenção é que conta.	139. Muito riso, pouco siso.	253. Quem feio ama bonito lhe parece.
15. A justiça tarda mas não falta.	140. Na hora do aperto é que se conhecem os amigos.	254. Quem foi ao mar, perdeu o lugar.
16. A laranja, de manhã é ouro, à tarde é prata e à noite mata.	141. Não deixes para amanhã o que podes fazer hoje.	255. Quem mais jura mais mente.
17. A mentira tem perna curta.	142. Não é com vinagre que se apanham moscas.	256. Quem muito dorme pouco aprende.
18. A montanha pariu um rato.	143. Não façás aos outros o que não queres que te façam a ti.	257. Quem muito escolhe pouco acerta.
19. A morte não escolhe idades.	144. Não há amor como o primeiro.	258. Quem muito fala, pouco acerta.
20. A mulher e a sardinha querem-se da mais pequenina.	145. Não há bela sem senão.	259. Quem não arrisca não petisca.
21. A noite é boa conselheira.	146. Não há bem que sempre dure nem mal que nunca acabe	260. Quem não chora não mama.
22. À noite todos os gatos são pardos.	147. Não há duas sem três.	261. Quem não deve não teme.
23. A ocasião faz o ladrão.	148. Não há regra sem exceção.	262. Quem não é por mim é contra mim.
24. A paciência tem limites.	149. Não há rosa sem espinhos.	263. Quem não é visto não é lembrado.
25. A pensar morreu um burro.	150. Não se fazem omeletes sem ovos.	264. Quem não quer ser lobo não lhe veste a pele.
26. A preguiça é a mãe de todos os vícios.	151. Não se pode ter tudo.	265. Quem não sabe é como quem não vê.
27. A pressa é inimiga da perfeição.	152. Nem oito nem oitenta.	266. Quem não tem cão caça com gato.
28. A roupa suja lava-se em casa.	153. Nem só de pão vive o homem.	267. Quem não tem dinheiro não tem vícios.
29. A saúde não tem preço.	154. Nem tanto ao mar, nem tanto à terra.	268. Quem não trabuca não manduca.
30. A união faz a força.	155. Nem tudo o que reluz é ouro.	269. Quem nasce torto tarde ou nunca se endireita.
31. A verdade vem sempre ao de cima.	156. Nem tudo o que vem à rede é peixe.	270. Quem pergunta quer saber.
32. A vida são dois dias.	157. Ninguém é de ferro.	271. Quem pode pode.
33. A vingança serve-se fria.	158. Ninguém nasce ensinado.	272. Quem procura acha.
34. A voz do povo é a voz de Deus.	159. No meio é que está a virtude.	273. Quem ri por último ri melhor.
35. Abril, águas mil.	160. No melhor pano cai a nódoa.	274. Quem sabe sabe.
36. Achado não é roubado.	161. No poupar é que está o ganho.	275. Quem sai aos seus não degenera.
37. Água mole em pedra dura tanto bate até que fura.	162. Nunca digas nunca.	276. Quem se mete em atalhos, mete-se em trabalhos.
38. Águas passadas não movem moinhos.	163. Nunca é tarde para aprender.	277. Quem semeia ventos colhe tempestades.
39. Albarde-se o burro à vontade do dono.	164. O amor é cego.	278. Quem semeia, colhe.
40. Amanhã é outro dia.	165. O ataque é a melhor defesa.	279. Quem te avisa teu amigo é.
41. Amanhã também é dia.	166. O barato sai caro.	280. Quem tem boca vai a Roma.
42. Amigo não empata amigo.	167. O bom filho à casa torna.	281. Quem tem cu tem medo.
43. Amigos, amigos, negócios à parte.	168. O crime não compensa.	282. Quem tem medo compra um cão.
44. Amor com amor se paga.	169. O dinheiro não traz felicidade.	283. Quem tem telhados de vidro não atira pedras ao do vizinho.
45. Ano novo, vida nova.	170. O fruto proibido é o mais apetecido.	284. Quem tem unhas é que toca guitarra.
46. Antes só que mal acompanhado.	171. O futuro a Deus pertence.	285. Quem tudo quer tudo perde.
47. Aprender até morrer.	172. O hábito não faz o monge.	286. Quem vai à guerra dá e leva.
48. As ações ficam com quem as pratica.	173. O meu é meu, o teu é nosso.	287. Quem vai para o mar avia-se primeiro em terra.
49. As aparências iludem.	174. O pior cego é o que não quer ver.	288. Quem vê caras não vê corações.
50. As árvores morrem de pé.	175. O primeiro milho é dos pardais.	289. Quem vier atrás que feche a porta.
51. As más notícias chegam depressa.	176. O prometido é devido.	290. Querer é poder.
52. As paredes têm ouvidos.	177. O que arde cura.	291. Recordar é viver.
53. Até ao lavar dos cestos é vindima.	178. O que é bom acaba depressa.	292. Rei morto, rei posto.
54. Azar ao jogo, sorte no amor.	179. O que é doce nunca amargou.	293. Rir é o melhor remédio.
55. Beleza não põe mesa.	180. O que lá vai lá vai.	294. Roma e Pavia não se fizeram num dia.
56. Boas contas fazem os bons amigos.	181. O que não mata, engorda.	295. Saber esperar é uma virtude.
57. Boda molhada, boda abençoada.	182. O que não tem remédio, remediado está.	296. Santos da casa não fazem milagres.
58. Burro velho não aprende línguas.	183. O que tem de ser tem muita força.	297. São mais as vozes que as nozes.
59. Cá se fazem, cá se pagam.	184. O saber não ocupa lugar.	298. Se Maomé não vai à montanha, a montanha vai a Maomé.
60. Cada cabeça sua sentença.	185. O segredo é a alma do negócio.	299. Sem trabalho nada se consegue.
61. Cada cavadela sua minhoca.	186. O seguro morreu de velho.	300. Só se vive uma vez.
62. Cada coisa a seu tempo.	187. O seu a seu dono.	301. Tal pai, tal filho.
63. Cada macaco no seu galho.	188. O silêncio é de ouro.	302. Todos os caminhos vão dar a Roma.
64. Cada maluco com a sua mania.	189. O sol quando nasce é para todos.	
65. Cada qual com o seu igual.	190. O tempo é dinheiro.	
66. Cada um é como cada qual.	191. O tempo tudo cura.	
67. Cada um é para o que nasce.		
68. Cada um por si e Deus por todos.		
69. Cada um puxa a brasa à sua sardinha.		
70. Cada um sabe as linhas com que se cose.		
71. Cada um tem aquilo que merece.		

72. Candeia que vai à frente alumia duas vezes.	192. O tempo voa.	503. Toma lá, dá cá.
73. Cão que ladra não morde.	193. Olho por olho, dente por dente.	504. Três foi a conta que Deus fez.
74. Casa onde não há pão, todos ralham e ninguém tem razão.	194. Olhos que não veem, coração que não sente.	505. Tristezas não pagam dívidas.
75. Casa roubada, trancas à porta.	195. Onde há fumo, há fogo.	506. Tudo está bem quando acaba bem.
76. Cautela e caldos de galinha nunca fizeram mal a ninguém.	196. Os amigos são para as ocasiões.	507. Um abismo chama outro abismo.
77. Com o fogo não se brinca.	197. Os cães ladram e a caravana passa.	508. Um burro carregado de livros é um doutor.
78. Com papas e bolos se enganam os tolos.	198. Os erros pagam-se caro.	509. Um crime não justifica outro.
79. Com uma mentira se apanha uma verdade.	199. Os fins justificam os meios.	510. Um homem prevenido vale por dois.
80. Contas são contas.	200. Os gostos não se discutem.	511. Um mal nunca vem só.
81. Contra factos não há argumentos.	201. Os homens não se medem aos palmos.	512. Uma imagem vale mais que mil palavras.
82. Cresce e aparece.	202. Os olhos comem mais que a barriga.	513. Uma mão lava a outra.
83. Da discussão nasce a luz.	203. Os olhos são o espelho da alma.	514. Vão-se os anéis, ficam os dedos.
84. De boas intenções está o inferno cheio.	204. Os opostos atraem-se.	515. Velhos são os trapos.
85. De Espanha nem bom vento, nem bom casamento.	205. Os últimos serão os primeiros.	516. Vivendo e aprendendo.
86. De pequenino é que se torce o pepino.	206. Ou oito ou oitenta.	517. Vozes de burro não chegam ao céu.
87. De poeta e de louco, todos temos um pouco.	207. Palavra puxa palavra.	518. Zangam-se as comadres, descobrem-se as verdades.
88. Deitar cedo e cedo erguer, dá saúde e faz crescer.	208. Palavras loucas, orelhas moucas.	
89. Depois da tempestade vem a bonança.	209. Palavras, leva-as o vento.	
90. Depressa e bem não há quem.	210. Pão pão, queijo queijo.	
91. Deus dá nozes a quem não tem dentes.	211. Para a frente é que é o caminho.	
92. Deus escreve direito por linhas tortas.	212. Para bom entendedor, meia palavra basta.	
93. Deus não dorme.	213. Para grandes males, grandes remédios.	
94. Devagar se vai ao longe.	214. Para morrer basta estar vivo.	
95. Dia de São Martinho, lume, castanhas e vinho.	215. Para quem é, bacalhau basta.	
96. Diz-me com quem andas, dir-te-ei quem és.	216. Parar é morrer.	
97. Do mal, o menos.	217. Patrão fora, dia santo na loja.	
98. Dos fracos não reza a História.	218. Peixe não puxa carroça.	
99. É mais fácil apanhar um mentiroso que um coxo.	219. Pela boca morre o peixe.	
100. É pior a emenda que o soneto.	220. Perdido por cem, perdido por mil.	
101. É preciso dar tempo ao tempo.	221. Por causa de uns, pagam os outros.	
102. Em boca fechada não entra mosca.	222. Por morrer uma andorinha não acaba a Primavera.	
103. Em casa de ferreiro, espeto de pau.	223. Porta da rua é serventia da casa.	
104. Em Roma sê romano.	224. Quando a cabeça não tem juízo, o corpo é que paga.	
105. Em tempo de guerra não se limpam armas.	225. Quando a esmola é grande, o pobre desconfia.	
106. Em terra de cegos quem tem olho é rei.	226. Quando o mar bate na rocha quem se lixa é o mexilhão.	
107. Enquanto há vida há esperança.	227. Quando um burro fala, os outros baixam as orelhas.	
108. Enquanto o pau vai e vem folgam as costas.	228. Quando uma porta se fecha outra se abre.	
109. Entre marido e mulher, nunca metas a colher.	229. Quanto mais alto se sobe, maior é a queda.	
110. Entre mortos e feridos alguém há-de escapar.	230. Quanto mais depressa mais devagar.	
111. Errar é humano.	231. Quanto mais se tem, mais se quer.	
112. Faz o que eu digo e não faças o que eu faço.	232. Quem anda à chuva molha se.	
113. Filho de peixe sabe nadar.	233. Quem boa cama fizer nela se deitará.	
114. Filhos criados, trabalhos dobrados.	234. Quem brinca com o fogo queima se.	
115. Gaivotas em terra, tempestade no mar.	235. Quem cala consente.	
116. Gato escaldado de água fria tem medo.	236. Quem canta seus males espanta.	
117. Gordura é formosura.	237. Quem casa quer casa.	
118. Grande nau, grande tormenta.	238. Quem com ferro fere com ferro será ferido.	
119. Grão a grão enche a galinha o papo.	239. Quem conta um conto acrescenta um ponto.	
120. Há gente para tudo.		
121. Há mais marés que marinheiros.		
122. Há males que vêm por bem.		
123. Há mar e mar, há ir e voltar.		
124. Há sempre uma primeira vez para tudo.		
125. Hoje por mim, amanhã por ti.		

Capítulo 10

As intenções comunicativas no uso de provérbios

Aceite para publicação em: Soares, R. & Lauhakangas, O. (Eds.) *13th Interdisciplinary Colloquium on Proverbs, Actas ICP19 Proceedings*. Tavira: AIP-IAP, 2019



AS INTENÇÕES COMUNICATIVAS NO USO DE PROVÉRBIOS

Sónia REIS¹ & Jorge BAPTISTA^{1,2}, ¹Universidade do Algarve, ²INESC-ID, *Portugal*

Resumo

O objetivo deste artigo é contribuir para a reflexão em torno da intenção comunicativa no uso de provérbios no discurso. À semelhança do que acontece com outros tipos de expressões, nem sempre é fácil determinar com que intenção é utilizado um provérbio numa determinada situação comunicativa. Embora seja possível construir definições ou descrições que procurem dar conta do significado genérico e do contexto de uso de muitos provérbios, um mesmo provérbio parece poder desempenhar funções pragmáticas diferentes consoante a situação em que é usado, o que faz variar subtilmente o seu significado em função da intenção comunicativa do locutor. Assim, a fim de determinar a intenção comunicativa do uso de provérbios no discurso, desenvolvemos e aplicámos um inquérito por questionário, que procura captar (ainda que parcialmente) esse conhecimento que o falante tem e, indiretamente, determinar a sua intenção comunicativa.

Palavras-chave: Funções pragmáticas, intenção comunicativa, provérbio.

1. Introdução

Os provérbios assumem as mais variadas funções discursivas nos contextos onde ocorrem. Ainda que surjam integrados no discurso, apresentam uma certa autonomia relativamente ao contexto em que são usados e, por si só, podem veicular uma variedade de significados, mais ou menos literais/convencionais.

A motivação para o uso de provérbios poderá depender de diversos fatores: a intenção do emissor; a facilidade da sua utilização, uma vez que são estruturas disponíveis, concisas e que são reconhecidas por uma comunidade linguística; a situação; o contexto comunicativo, ou outras.

De qualquer forma, existe um conjunto de modalidades típicas associadas aos provérbios. Por exemplo, Lopes (1992), relativamente à modalidade, distingue duas classes básicas de provérbios – os *provérbios descritivos*, com um valor ilocutório *assertivo*, e.g. *Cada santo tem seu nicho*; e os *provérbios normativos*, com um valor ilocutório *diretivo*, e.g. *Quando fores bigorna, sofre, e quando fores malho, malha*.

À semelhança do que acontece com outros elementos da língua, nem sempre é fácil determinar com que intenção comunicativa é utilizado um determinado provérbio. Embora seja possível construir definições ou descrições que procurem dar conta do significado genérico e do contexto de uso de muitos provérbios, um mesmo provérbio pode desempenhar

funções pragmáticas diferentes consoante a situação em que é usado, o que faz variar subtilmente o seu significado em função da intenção comunicativa do locutor.

Efetivamente, os provérbios assumem variadas funções pragmáticas nos textos em que ocorrem: Podem ter como finalidade **aconselhar** (e.g. *Quando a Lua minguar, não debes semear*); **advertir/avisar** (e.g. *Cuidado com o homem calado e com o cão que não ladra*); **instruir/informar** (e.g. *Março marçação, manhã de inverno, tarde de verão*); **ordenar** (e.g. *Come, que o teu mal é fome*); **pedir** (e.g. *Ajuda-me que eu te ajudarei*); entre outras (Faria, I., 2003, Cap. 4; Mendes, A., 2013, Vol. II, Cap. 34; Raposo, E., 2013, Vol. I, Cap. 11).

Nem sempre é fácil distinguir estes diferentes valores de modo reprodutível, sendo difícil determinar um conjunto de categorias pragmáticas fundamentais, relevantes e consensuais, para descrever as intenções comunicativas subjacentes à utilização dos provérbios em discurso. Partimos, por isso, do princípio de que se o falante conhece e sabe usar um provérbio, também conhece e tem interiorizadas essas categorias, ainda que não seja tarefa fácil enunciá-las.

Assim, a fim de determinar a intenção comunicativa do uso de provérbios no discurso, concebemos duas experiências, que procuram captar (ainda que parcialmente) esse conhecimento intuitivo do falante e, indiretamente, indicar a sua intenção comunicativa quando usa um dado provérbio. Descreveremos estas experiências mais adiante neste trabalho.

Duas questões orientam esta investigação:

- 1) Será possível determinar/aférir com que intenção comunicativa é utilizado um determinado provérbio?
- 2) Até que ponto é possível determinar, de modo rigoroso e consensual, a função pragmática dos provérbios, atendendo às intenções dos falantes e às relações que os provérbios podem estabelecer com o contexto textual em que são utilizados?

Tendo estes aspetos em consideração, pretendemos: (i) contribuir para o estudo das intenções comunicativas no uso de provérbios quando integrados em discurso e (ii) identificar as diferentes funções pragmáticas que os provérbios podem desempenhar de acordo com a situação em que são utilizados.

Os resultados obtidos poderão contribuir para um melhor conhecimento das intenções comunicativas associadas ao uso dos provérbios, bem como uma reflexão sobre a perceção que os falantes têm quanto às condições pragmáticas que autorizam o uso de um provérbio no discurso.

2. Trabalhos relacionados

Neste ponto do trabalho, apresentamos alguns estudos de provérbios portugueses em que se analisam quer as relações que as orações por que são constituídos os provérbios estabelecem entre si, quer as relações que os provérbios estabelecem com os contextos discursivos em que se encontram.

Meira (2011), partindo da Rhetorical Structure Theory (RST) e das proposições relacionais¹ (Mann & Thompson, 1983), desenvolveu um estudo sobre a articulação de orações em 20 provérbios do português do Brasil, analisando as relações retóricas que se observam no interior dos mesmos. A autora adota a teoria funcionalista (Halliday, 2004), considerando que uma teoria que se limitasse ao nível da frase se tornaria limitadora por não considerar critérios pragmáticos. Esta teoria descreve também as relações implícitas que se encontram entre as orações que se articulam nos provérbios. A autora procura, pois, observar a competência comunicativa dos falantes na codificação e descodificação dos provérbios e averiguar a sua capacidade para interpretar esses provérbios numa situação de uso. As relações retóricas são definidas com base na RST. A partir das proposições relacionais existentes na articulação entre frases, determinaram-se quais as relações retóricas que predominam numa dada situação comunicativa e que são, na sua maioria, relações de *circunstância*, *elaboração*, *contraste*, *condição* e *concessão*. A autora refere ainda que o contexto discursivo teve um papel decisivo para que essas relações fossem compreendidas.

Num outro trabalho (Meira, 2015), ainda sobre as funções retóricas dos provérbios, a mesma autora faz um estudo cujo *corpus* de trabalho é constituído por 10 provérbios sem núcleo verbal e sem conetores (*e.g. Casa de ferreiro, espeto de pau*), analisados em 20 textos (cada provérbio foi analisado em dois contextos distintos), com o objetivo de identificar o papel das relações retóricas na coerência e na organização textual². A autora constatou que certas informações (genéricas e específicas) dos provérbios permitem que sejam validadas as relações tidas como possíveis para um dado provérbio. Meira verificou ainda que as relações retóricas dos provérbios não se alteraram em 70% dos casos, em função de uma situação específica. Ainda assim, a autora refere que o contexto tem grande importância para reafirmar a escolha de uma determinada relação retórica, uma vez que a situação específica apresenta aspetos que contribuem para a justificação da relação retórica determinada.

Dias 2009, ao estudar os tipos de “combinação de cláusulas”, faz uma análise das expressões proverbiais justapostas, em que se observa a relação “adverbial implícita” de estruturas como em *Casa de ferreiro, espeto de pau*. O seu *corpus* de trabalho é constituído por 20 textos que

apresentam provérbios no título, quer sob as “suas fórmulas fixas originais”, quer sob “re-escrituras delas” (*idem*: 17). A autora refere que, por meio de paráfrases, é possível observar as relações implícitas nas frases, no caso particular do provérbio referido a relação de *concessão* (“Embora a casa seja de ferreiro, o espeto é de pau” (*idem*: 42). Opinião diferente tem Meira (2013) que defende que não adianta apresentar paráfrases para estabelecer uma equivalência entre frases nominais e orações, mas sim demonstrar que qualquer uma destas pode figurar como um texto, daí as definições tradicionais não serem satisfatórias. Voltando a Dias (2009) é graças a estas paráfrases entre as partes das construções proverbiais justapostas que é estabelecida uma articulação hipotática. Relata também que a articulação hipotática difere da *subordinação*, uma vez que esta última presume uma dependência total de uma “cláusula” relativamente à outra, enquanto que na articulação hipotática há interdependência entre as “cláusulas”. Refere ainda que é necessário ter-se em consideração o contexto de uso das estruturas proverbiais para poder determinar a relação implícita existente entre as suas frases constituintes.

Lopes (1998) analisa um conjunto de provérbios inseridos em textos de imprensa, com o objetivo de contribuir para o estudo das funções discursivas dos provérbios. A autora refere que para se perceber o funcionamento pragmático típico do provérbio, é necessário analisar o uso efetivo do provérbio inserido em processos variados de interação verbal. Verificou que os provérbios que ocorrem no seu *corpus* de trabalho aparecem em diferentes lugares: (i) em posição de destaque (por exemplo em títulos); (ii) no interior de *leads*; (iii) inseridos no corpo de artigos ou crónicas, às vezes funcionando como remate conclusivo. Constatou que os provérbios, independentemente da posição em que surgem, podem ser *citados* na íntegra, parcialmente *truncados* ou mesmo *distorcidos*. Do ponto de vista pragmático, a autora identificou as seguintes funções discursivas dos provérbios: (a) em títulos e em subtítulos, a função de resumo catafórico daquilo que será explicitado depois. Nestes casos, ou desempenhando um papel “retórico” ou correspondendo a um comentário (de forma indireta) do próprio jornalista; (b) no final de artigos ou crónicas, a função de remate/comentário ou *sumário anafórico* do texto; e (c) no interior de artigos e crónicas, a função de *comentário avaliativo* (do jornalista) ou *como peça de um esquema argumentativo*. Neste último caso, é geralmente argumento a favor de uma conclusão ou mesmo justificação de uma afirmação anterior. Em forma de síntese, a autora agrupa estas funções em duas classes distintas: 1) “as que relevam de uma análise das funções ilocutórias dos provérbios inseridos em sequência discursivas” [...] 2) “as que relevam de um plano de organização macro-textual” (*idem*: 125). Relativamente à distorção dos provérbios, a autora afirma que esta distorção tanto poderá ter

uma intenção lúdica como poderá ter uma intenção subversiva, quando parece *desconstruir* a sabedoria popular.

Ainda sobre os usos e as intenções retóricas dos provérbios em textos da imprensa, também Figueiredo (2012) apresenta um estudo nesse sentido. A autora pretende verificar de que forma a escolha de um provérbio por um jornalista se relaciona com os seus propósitos comunicativos. Para tal, analisa um *corpus* de trabalho composto por quatro jornais que circularam ao longo dos séculos XIX e XXI no estado do Pará (Brasil). A autora verificou que os autores dos textos usavam os provérbios, em grande parte, para apresentar pontos de vista, aconselhar, ensinar, persuadir e coagir os leitores. A autora verificou também que, nos jornais mais antigos do seu *corpus* de trabalho, a escrita era mais espontânea e que a utilização de expressões populares, nomeadamente provérbios, era muito superior à dos períodos subsequentes. Acrescenta ainda que, nessa época, o comentário do próprio jornalista era mais importante do que a notícia em si.

Figueiredo refere que os jornais da atualidade, por sua vez, apresentam uma certa “desproverbialização” dos próprios provérbios ou assumem mesmo uma *cultura antiproverbial*, muitas vezes com a intenção de provocar o riso, satirizar hábitos e costumes, e questionar as verdades tidas como “infalíveis” e “imutáveis”.

A autora conclui que os provérbios assumem diversas funções retóricas, que são reguladas pela relação entre o contexto de uso dos provérbios, a época e que foram usados e a pessoa que os utiliza.

A partir destes trabalhos pudemos constatar que o contexto/situação de uso assume, efetivamente, uma papel importante para se poder determinar as intenções comunicativas que estão implícitas na utilização do provérbio.

3. Metodologia

A fim de determinar a intenção comunicativa do uso de provérbios quando integrados em discurso, desenvolvemos um inquérito por questionário que procura captar esse conhecimento que o falante tem e, indiretamente, determinar a sua intenção comunicativa.

Para o questionário, selecionámos aleatoriamente 10 provérbios, a partir de uma lista de 100 provérbios muito usuais, cuja disponibilidade lexical já foi atestada em estudos anteriores (Reis & Baptista, 2017; Reis et al. 2017). Esta escolha procurou garantir que todos os informantes conheçam bem e usavam efetivamente os provérbios selecionados. Depois, recorrendo à internet, extraímos, também de forma aleatória, 3 exemplos diferentes de uso de

cada um desses provérbios, o que resultou em 30 textos diferentes. A seleção destes textos assentou num conjunto de critérios.

Foram selecionados os três primeiros textos que apresentavam o provérbio-alvo e excluídos os casos que diziam respeito a:

- i. listas e coletâneas de provérbios (uma vez que não é possível estabelecer qualquer relação retórica entre provérbios e as listas em que se encontram);
- ii. títulos (pois ainda que possam aparecer muitas vezes como mote daquilo que se vai falar, não se encontram tão diretamente integrados no discurso como os que ocorrem no corpo do texto);
- iii. traduções (pelos eventuais problemas ou interferências que possam surgir entre a língua de partida e a língua de chegada);
- iv. monografias, teses e outros estudos paremiológicos (já que neles aparecem os provérbios que são objeto da discussão e não os provérbios integrados no discurso).

Dos textos selecionados, foram usados excertos com o máximo de 150 palavras, excluindo o próprio provérbio. Nos casos em que o número de palavras era inferior a 150 foi utilizado todo o texto disponível. Sempre que possível, foram selecionadas apenas as 75 palavras imediatamente antes e depois de cada provérbio, para dar conta do seu contexto.

O questionário apresenta uma explicação introdutória, onde são indicados os objetivos e instruções. São ainda recolhidas algumas informações pessoais dos respondentes, nomeadamente, o género, a idade, a escolaridade, a nacionalidade, o distrito e a sua área de residência. Os provérbios foram apresentados sempre na mesma ordem a todos os inquiridos para que as condições de aplicação fossem as mesmas para todos os respondentes.

Trata-se de um questionário que exige uma grande capacidade de concentração, uma vez que é pedida aos inquiridos a realização de diferentes tarefas. Estas tarefas consistem em:

- i. leitura do texto com o provérbio-alvo
- ii. identificação dos elementos ou expressões sobre os quais é dito o provérbio;
- iii. seleção, cópia a colagem desses elementos para o local destinado para o efeito; e ainda

- iv. seleção de entre uma das opções de escolha múltipla apresentadas.

Num primeiro momento era pedido aos respondentes que lessem o texto apresentado, onde se encontra o provérbio-alvo, que foi assinalado entre cardinais, de forma a facilitar a sua visualização. Após a identificação dos elementos ou expressões sobre os quais o provérbio é dito, o respondente deveria selecionar, copiar e colar esses elementos para o local destinado para o efeito. A tarefa seguinte consistia na seleção de uma das opções de escolha múltipla apresentadas que pudesse determinar a polaridade (positiva-neutra-negativa) do provérbio.

Foi usada uma escala de Likert de 5 níveis, e ainda apresentamos as opções: *não sei o que responder e não é possível, com base no excerto apresentado, determinar a que se refere o provérbio.*

Para podermos analisar os dados obtidos através da aplicação deste questionário foi necessário estabelecer uma referência. Para tal, dois especialistas, independentemente, identificaram o tópico a propósito do qual o provérbio era dito, e delimitaram os segmentos do texto que denotavam esse mesmo tópico. Depois, foi necessário tentar determinar a intenção comunicativa subjacente, já que a polaridade (positiva ou negativa) de uma afirmação se relaciona com a perceção que o informante desenvolve quanto à atitude do autor do texto relativamente ao tópico sobre o qual o provérbio tinha sido dito. Após confrontados os resultados dos especialistas, foi estabelecida uma anotação de referência que serviu para analisar os dados.

O inquérito foi disponibilizado no dia 19 de novembro de 2018 e obteve a última resposta em janeiro de 2019. Selecionou-se uma amostra por conveniência, recorrendo a contactos informais para a seleção dos respondentes. Este inquérito foi então distribuído através de uma lista de cerca de 200 contactos pessoais.

No ponto seguinte deste trabalho, mostraremos os resultados obtidos da aplicação do questionário.

4. Resultados

Da aplicação do questionário *As intenções comunicativas no uso de provérbios*, obteve-se 34 respostas. A taxa de resposta é de 17%.

Os sujeitos da amostra são maioritariamente do sexo feminino (71%), atestando uma grande assimetria de género dos respondentes. 71% dos inquiridos têm idades compreendidas entre os 31 e os 50 anos e somente 6% têm menos de 31 anos. De acordo com os dados recolhidos, 65% possuem o ensino superior e os restantes o ensino secundário. Os inquiridos são todos

portugueses, tendo um destes dupla nacionalidade. Relativamente à área de residência, 76,5% vivem em zona urbana e 23,5% em zona rural, demonstrando uma grande assimetria da natureza da zona. Faro é o distrito de residência em que se obteve o maior número de respostas, com 76,5% das respostas.

Os informantes reportaram que foi necessário muito tempo para responder ao questionário (cujo tempo previsto era de 30 minutos). Referem ainda que tiveram dificuldade em identificar o tópico sobre o qual o provérbio era dito e em atribuir um valor no eixo positivo/negativo.

De forma a simplificar a apresentação dos resultados obtidos no questionário, iremos agrupar os dados associando-os a duas questões diferentes – *Questão 1* e *Questão 2*. Assim a *Questão 1* reúne os dados referentes às tarefas que consistiam (i) na leitura de textos com o provérbio-alvo (ii) na identificação dos elementos ou expressões sobre os quais é dito o provérbio (iii) na seleção, cópia a colagem desses elementos para o local destinado para o efeito; e a *Questão 2* concerne aos dados recebidos (iv) da seleção de entre uma das opções de escolha múltipla apresentadas em relação à polaridade atribuída aos provérbios. As opções de escolha múltipla não só apresentavam 5 graus diferentes de polaridade (*positivo; mais positivo do que negativo; nem positivo nem negativo; mais negativo do que positivo; negativo*), como também apresentavam as opções *não sei o que responder* e *não é possível, com base no excerto apresentado, determinar a que se refere o provérbio*.

Nas Figura 1 e Figura 2, apresentam-se os resultados referentes à *Questão 1*.

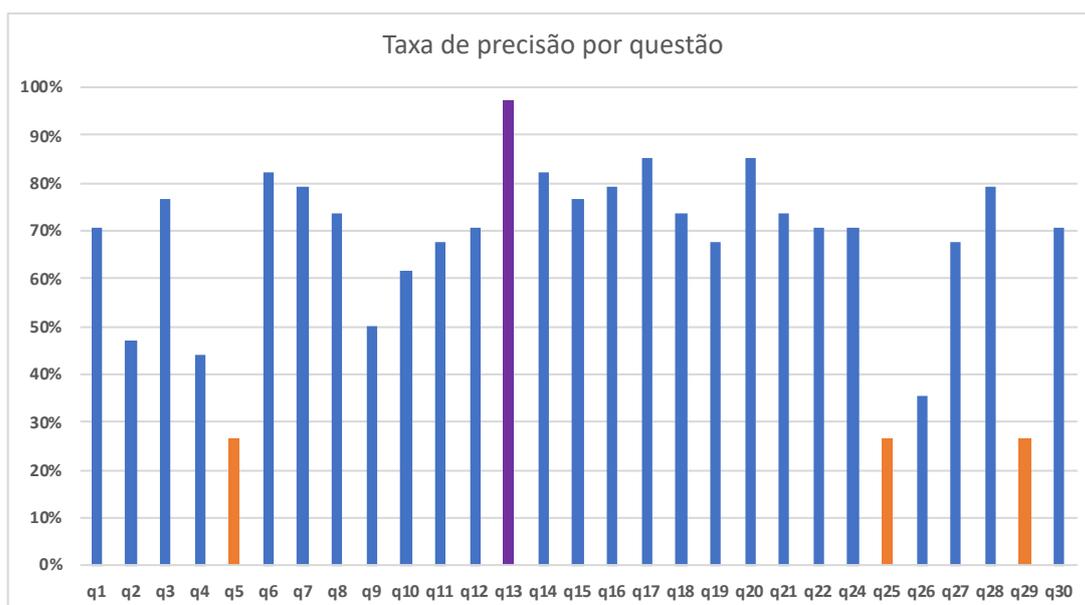


Figura 1: Taxa de precisão por questão (%) - Questão 1

Os dados apresentados na Figura 1 mostram que em média, 66% das respostas selecionadas pelos 34 respondentes correspondem à referência estabelecida pelos 2 anotadores. Ou seja, relativamente à *Questão 1* em que os respondentes deveriam identificar o tópico a propósito do qual o provérbio era dito, e delimitar os segmentos do texto que denotavam esse mesmo tópico, uma grande parte dos respondentes concorda com os anotadores relativamente a esses elementos.

Verificámos ainda que há uma grande variação da taxa de acerto entre questões. No caso da *q13*, 97% das pessoas acertaram na resposta. Esta resposta representa o 1º contexto de uso do provérbio *O fruto proibido é o mais apetecido*. Há, no entanto casos em que a taxa de acerto por resposta é muito baixa, nomeadamente as respostas às questões *q5*, *q25* e *q29*, em que a taxa de acerto é inferior a 30%.

No gráfico seguinte (Figura 2), apresenta-se a taxa de precisão por pessoa.

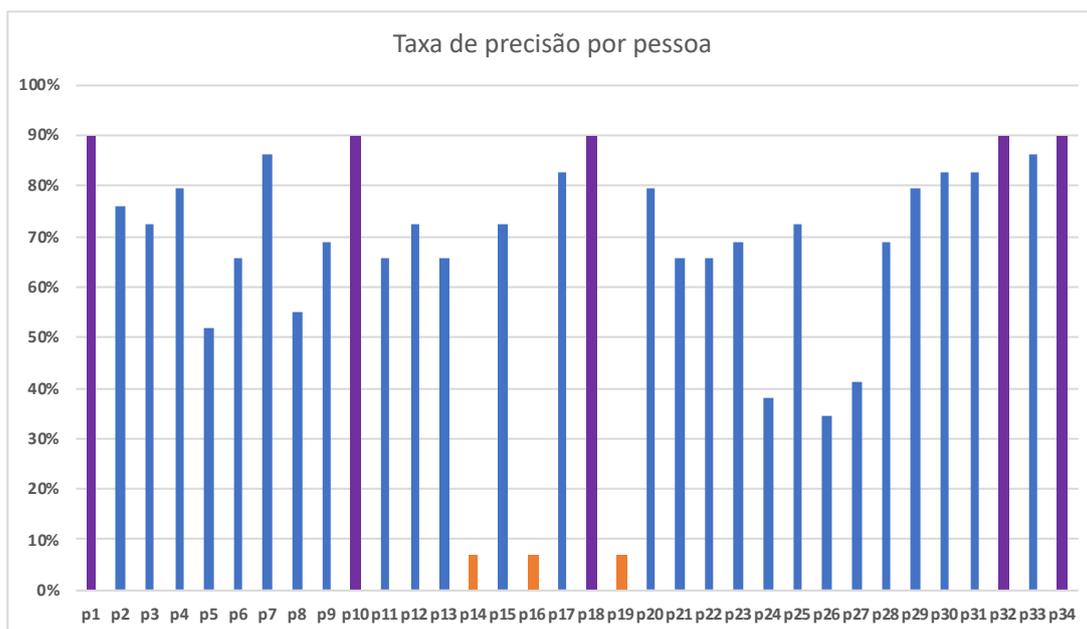


Figura 2: Taxa de precisão por pessoa (%) – Questão 1

Em média, 66% das pessoas acertaram na resposta certa. Há, no entanto, uma grande variação da taxa de acerto entre as pessoas. Destacam-se as cinco pessoas que acertaram em 90% das respostas, em *p1*, *p10* e *p18*, *p32* e *p34*, e as três pessoas que acertaram em apenas 7% das respostas, em *p14*, *p16* e *p19*.

Nas figuras seguintes (Figura 3 e Figura 4), são mostrados os dados relativos à *Questão 2*.

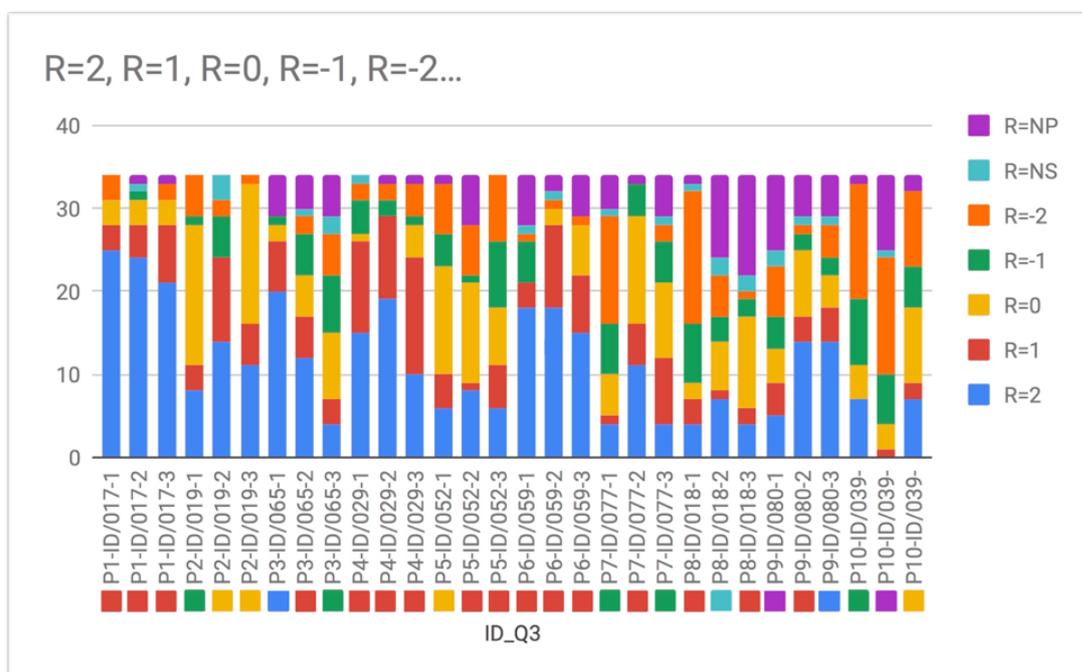


Figura 3: *Dispersão Respostas (N=34) Questão 2-polaridade*

Neste gráfico apresentamos a dispersão das 34 respostas obtidas, no que respeita à questão da polaridade.

À direita, na vertical, mostramos a legenda. Começando por baixo, a azul, R=2 equivale às respostas em que foi atribuído um valor positivo àquilo que o provérbio dizia sobre os elementos do texto; a vermelho, R1 designa o valor mais positivo do que negativo; a amarelo, R0 designa nem positivo nem negativo; a verde, R-1 respeita ao valor mais negativo do que positivo e a laranja R=-2 corresponde ao valor negativo. Temos ainda NS e NP, a azul turquesa e magenta, que representam, respetivamente, as opções *não sei o que responder* e *não é possível, com base no excerto apresentado, determinar a que se refere o provérbio*.

Em baixo, na horizontal, apresentamos o código atribuído a cada provérbio em cada um dos 3 contextos de uso, e os quadrados mais abaixo são referentes à referência estabelecida pelos anotadores.

Como se pode verificar, houve uma grande dispersão de respostas e nem sempre a resposta mais votada pelos inquiridos corresponde à referência dos anotadores.

As respostas em que houve maior consenso entre os 34 respondentes estão representadas nas 3 primeiras colunas deste gráfico e correspondem aos 3 contextos de uso do provérbios *A vida são dois dias*.

A resposta em que houve menor consenso entre os 34 respondentes está representada na 9ª coluna e é referente ao provérbio *Os olhos comem mais que a barriga*, no 3º contexto de uso. Neste caso, para além de terem sido selecionadas as 7 opções de resposta, a opção que foi mais votada apresenta apenas 23,5% das respostas (8 respostas).

No que respeita aos dados da concordância entre a referência dos 2 anotadores e os 34 participantes, verifica-se que a resposta em que houve maior acordo entre os 2 anotadores e os 34 participantes está representada na coluna 7 e respeita ao provérbio *Os olhos comem mais que a barriga*. A resposta onde houve menor consenso está representada na coluna 5 e é referentes ao provérbio *Abril, águas mil*, no contexto 2. Aqui os anotadores consideraram que não era possível determinar a polaridade do provérbio, que tinha nesse contexto uma função meramente informativa e os respondentes selecionaram maioritariamente as respostas do eixo positivo (R=2 e R=1).

Na figura seguinte (Figura 4) mostramos os mesmos dados que foram aqui apresentados na Figura 3, mas representados de uma outra forma, mais abrangente.

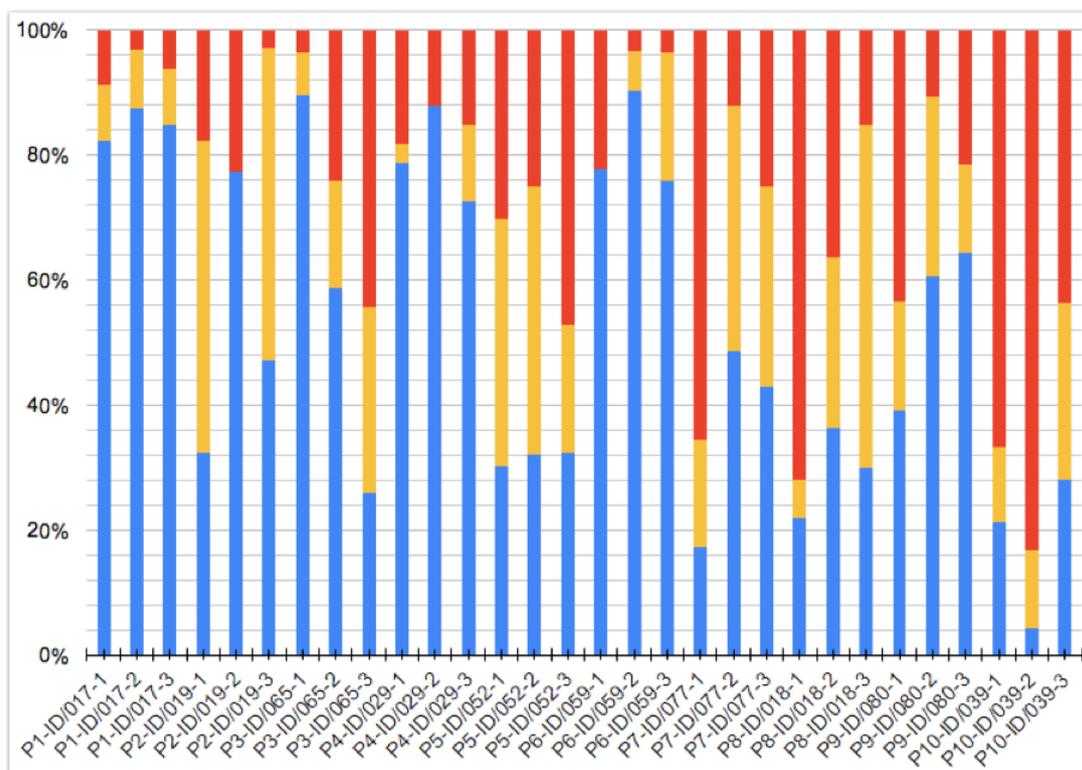


Figura 4: Normalização dados 34 anotadores (%) - *Positivo* [2, 1], *Neutro* [0], *Negativo* [-1, -2]

Este gráfico representa os mesmos dados que apresentámos no gráfico anterior, mas já normalizados. Retirámos as respostas referentes às opções *não sei o que responder* e *não é possível, com base no excerto apresentado, determinar a que se refere o provérbio* e organizámos os dados em três categorias apenas (positivo/neutro/negativo), juntando os valores que estavam mais próximos. O valor positivo está representado aqui a azul, o neutro a amarelo e o negativo a vermelho.

O resultado demonstra que parece haver uma predominância do valor positivo sobre o negativo. Há apenas 8 ocasiões em que o valor negativo é superior ao positivo. Destacam-se, ainda, vários casos em que a escolha do valor neutro é representativa, o que poderá significar que as pessoas tiveram dificuldade em atribuir um valor àquilo que o provérbios diziam sobre os elementos que seleccionaram do texto.

O grau de concordância entre os anotadores é mostrado a seguir, na Figura 5.

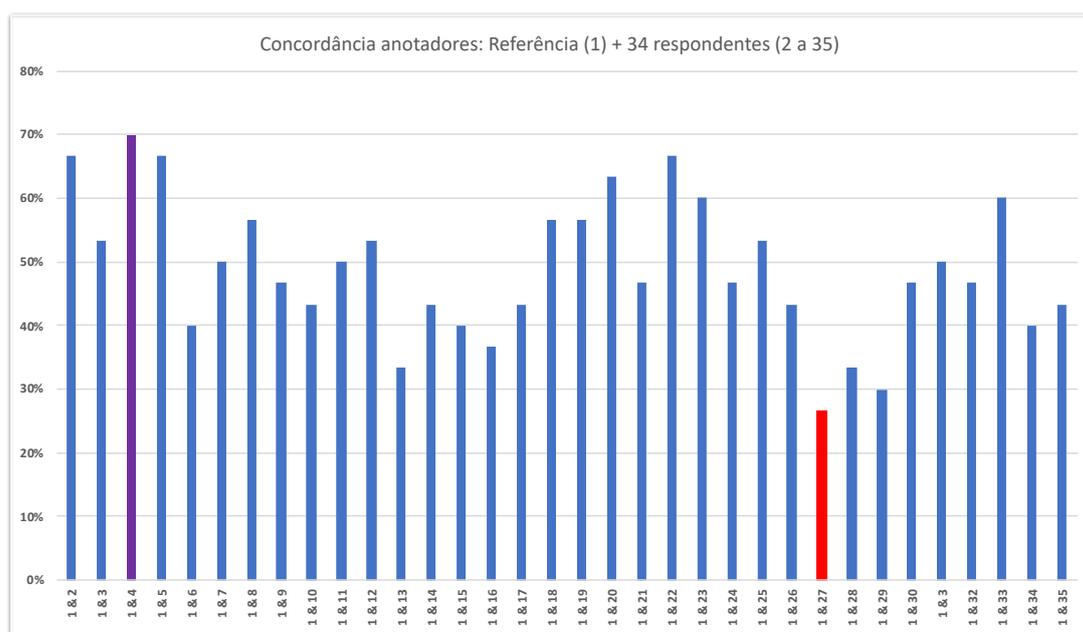


Figura 5: Acordo Referência 1 e 34 anotadores (%) Questão 2

Neste gráfico está representado o grau de concordância entre os anotadores Referência (em 1) e os anotadores Respondentes (de 2 a 35) ainda no que respeita à questão da polaridade.

Como se pode verificar, a taxa de acordo mais alta é entre o anotador referência e o anotador 4, que concordam 70% das vezes. Em média a referência e os anotadores concordam 49% das vezes. Apenas os pares 1/27 apresentam uma taxa de acordo inferior a 30%.

Conclusões e trabalho futuro

Com este estudo concluímos que há uma maior facilidade em identificar o tópico sobre o qual o provérbio é dito do que em atribuir um valor de polaridade àquilo que o provérbio diz.

Verificou-se que 66% das pessoas identificam o tópico do provérbio relativamente ao seu contexto e que 49% das pessoas determinam corretamente a polaridade (simplificada a 3 níveis) do provérbio. O que pode ser indicador de que o mesmo provérbio pode desempenhar diferentes funções pragmáticas no discurso. A função precisa de um provérbio dependerá (mais) do seu contexto e nem sempre pode ser ancorada no próprio provérbio.

É possível ainda concluir que a dificuldade na atribuição de um valor de polaridade não resulta da maior ou menor familiaridade do respondente com o provérbio, já que todos os provérbios aqui usados são *usuais* ou mesmo *muito usuais*.

A variação observada entre anotadores quanto à polaridade de um provérbio no mesmo contexto, bem como o número de provérbios em que a polaridade neutra foi a escolha predominante (5) parecem indicar que os anotadores não são capazes de atribuir, de forma consistente e reprodutível, um valor de polaridade a cada provérbio e até a um dado provérbio no mesmo contexto de uso. Tal parece tornar particularmente duvidoso que seja possível determinar objetivamente a função pragmática de um provérbio, já que na tarefa relativamente simples de atribuição de polaridade se observa tanta variação de respostas.

Naturalmente, estas conclusões são provisórias, podendo resultar de uma dificuldade do próprio estudo, nomeadamente o número de repostas limitado por amostragem de conveniência; a amostra de informantes relativamente enviesada (uma vez que a maioria dos respondentes são do sexo feminino; têm entre 31 e 50 anos; possuem as mesmas habilitações literárias; vivem em meio urbano e pertencem ao mesmo distrito). Poderá ainda dever-se à inadequação da escala de polaridade utilizada, entre outras.

Uma outra questão que deverá ser levantada é se será possível reconhecer as intenções comunicativas no uso de provérbios, e reconhecer fenómenos tão subtis como a ironia ou a vagueza.

No futuro, com base no questionário deste estudo, pretendemos desenvolver um outro questionário e aplicá-lo a uma amostra representativa da população, e, de acordo com os resultados obtidos, gostaríamos de contribuir para a construção de jogos didáticos assentes em relações semânticas e direcionados ao ensino de Português Língua não Materna (PLNM). O principal objetivo desses jogos seria tentar encontrar o provérbio apropriado para determinado contexto, com base em contextos de usos reais dos provérbios extraídos da internet.

Esta investigação foi parcialmente suportada por fundos nacionais através da Fundação para a Ciência e a Tecnologia (ref. UIDB/50021/2020).

Notas

1. As proposições relacionais podem ser expressas sem qualquer marca formal, uma vez que surgem da estrutura retórica do texto. Segundo Meira (2011: 39) [...] “podem ser definidas como o significado que emerge da combinação entre duas ou mais partes do texto, não sendo preciso que sejam partes adjacentes.” Por exemplo, a proposição relacional solução é definida como “uma parte do texto apresenta uma solução para o problema suscitado pela outra parte.” O exemplo da autora é: “Eu estou com fome. Vamos ao Fuji Gardens”. Trata-se, neste caso, de uma proposição implícita, isto é, que surge da relação entre as porções do texto mas não está lexicalmente expressa.
2. Critérios de seleção de provérbios e textos

Referências

- DIAS, Maria. 2009. A articulação hipotática em construções proverbiais justapostas. Dissertação de mestrado, Rio de Janeiro: Faculdade de Letras, Universidade Federal do Rio de Janeiro.
- FARIA, Isabel. 2003. O uso da língua, interação verbal e texto. In MATEUS et al., 2003. Gramática da Língua Portuguesa. 5ª edição revista e aumentada. Lisboa: Caminho., 55–84.
- FIGUEIREDO, Glauco. 2012. O gênero proverbial na imprensa: usos e funções retóricas. Tese de doutoramento, Recife: Universidade Federal de Pernambuco.
- HALLIDAY, Michael. 2004. *An Introduction to Functional Grammar*. 3ª edição. Londres, Hodder Education.
- LOPES, Ana. 1992. Texto Proverbial Português – Elementos para uma análise semântica e pragmática, Tese de Doutoramento, Coimbra: Faculdade de Letras da Universidade de Coimbra.
- LOPES, Ana. 1998. Funções discursivas do provérbio em textos de imprensa. Estudos de Literatura Oral, ELO, IV, 115-126.
- MANN, William & THOMPSON, Sandra. 1983. Relational proposition in discourse. ISI: Information Sciences Institute of University of Southern California, Los Angeles, ISI/RR-83-115.
- MEIRA, Ana. 2011. A articulação de orações em provérbios do português em uso: uma análise das relações retóricas. Dissertação de mestrado, Belo Horizonte: Universidade Federal de Minas Gerais.
- MEIRA, Ana. 2013. Amigos, amigos, negócios à parte: uma análise da estrutura retórica. *Anais do SILEL*. Volume 3, Número 1. Uberlândia: EDUFU.
- MEIRA, Ana. 2015. Casa de ferreiro, espeto de pau: uma análise das relações retóricas a partir do uso dos provérbios como estratégia argumentativa em textos da internet. Tese de doutoramento, Belo Horizonte: Universidade Federal de Minas Gerais.
- MENDES, Amália, 2013. Organização textual e articulação de orações. In RAPOSO, Eduardo Buzaglo Paiva et al. (Org.). Gramática do Português. Vol. II. Lisboa: Calouste Gulbenkian.
- RAPOSO, Eduardo, 2013. Estrutura da frase. In RAPOSO, Eduardo Buzaglo Paiva et al. (Org.). Gramática do Português. Vol. I. Lisboa: Calouste Gulbenkian.
- REIS, Sónia & BAPTISTA Jorge, 2017. Estimating Lexical Availability of European Portuguese Proverbs. In: Mitkov R. (eds) Computational and Corpus-Based Phraseology. EUROPHRAS 2017. Lecture Notes in Computer Science, vol 10596. Springer, Cham, 232-244.
- REIS, Sónia, POMPILI, Annamaria, ABAD, Alberto, BAPTISTA, Jorge, 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).

Curriculum vitae

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 é doutor em Linguística, na especialidade de Sintaxe, e professor associado com ‘tenure’ 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 recursos linguísticos e de aplicações computacionais para o ensino da língua.

Email: reis_soniamm@gmail.com, jbaptis@ualg.pt

Parte III

Conclusão

Chegados a este ponto do percurso, é altura de fazer, em jeito de *Conclusão*, uma avaliação do trabalho de investigação realizado, sintetizar os resultados alcançados e comentar os objetivos atingidos. Na nossa perspetiva, porém, esta tese não é a estação terminal deste percurso, mas sim um ponto de chegada e de partida para novas questões e novos estudos. Assim, num primeiro momento, faremos uma breve recapitulação dos problemas que nos propusemos abordar para, em seguida, fazermos uma síntese de resultados e sua discussão. Num segundo momento desta *Conclusão*, determinamos algumas linhas de investigação e pistas de trabalho futuro, que se abrem, no curto e no médio prazo, à investigação sobre as expressões proverbiais.

Partimos para esta tese com um objetivo geral, o de contribuir para um melhor conhecimento dos provérbios do português europeu. Ao longo da *Introdução*, levantámos algumas das problemáticas que considerámos as mais relevantes para descrição gramatical dos provérbios, sobretudo numa perspetiva lexicológica, e passámos em revista os aspetos teóricos que considerámos os mais relevantes nesta perspetiva. Procurando não nos repetir demasiado, recuperemos, nesta *Conclusão*, e de uma forma sucinta, os tópicos abordados.

Um dos aspetos problemáticos dos estudos paremiológicos prende-se com o conceito de provérbio enquanto unidade de descrição linguística: inclinamo-nos para as posições de outros autores, que incluem os provérbios, de forma mais ou menos implícita, no léxico da língua, já que se trata de um conhecimento partilhado numa comunidade linguística, à semelhança do que sucede com outro tipo de unidades lexicais, de natureza frásica (as expressões fixas/idiomáticas/cristalizadas) ou subfrásica (locuções ou palavras compostas). Não obstante, a sua natureza textual, nomeadamente, o facto de constituírem frases completas, mais ou menos longas, eventualmente truncadas, está estreitamente ligada às condições pragmáticas que presidem à sua utilização no discurso, condições que têm igualmente de ser aprendidas, sobretudo quando o significado global do provérbio ou o vocabulário envolvido não é suficiente para as inferir. O provérbio situa-se, assim, no extremo polo de uma conceção de unidade lexical abrangente, que, do ponto de vista formal, vai da palavra simples à frase completa

e, neste ponto, envolve uma irreduzível dimensão pragmática, muitas vezes ignorada na descrição lexical. Não se estranha, pois, que seja excluído dos dicionários de língua usuais e remetido, com as expressões idiomáticas, para o género específico do dicionário de provérbios, onde o seu estatuto lexical é, talvez, menos evidente. O provérbio é, de facto, um objeto linguístico incómodo, desconfortável, difícil de abordar, problemático para descrever, e, sobretudo, resiliente, pela sua diversidade estrutural, variação formal e extensão lexical, a generalizações redutoras.

Falámos, nessa sequência, das complexas relações que se estabelecem entre provérbios e expressões fixas ou idiomáticas. Nesse âmbito, recuperámos o conceito de *idiomaticidade*, entendido aqui exclusivamente no sentido de *não-composicionalidade semântica*, e reiterámos as posições de alguns autores, para os quais tal não é um critério absoluto e necessário para a determinação do estatuto de provérbio destas expressões. Efetivamente, em *Não faças aos outros o que não gostavas que te fizessem a ti*, é difícil detetar alguma leitura não literal para qualquer dos elementos da expressão. Pelo contrário, a expressão *Muita parra, pouca uva* aplica-se a “alguém que fala muito mas [que], no fundo, não diz nada que se aproveite”¹ mas este significado (e outros semelhantes com que possa ser relacionada), nada tem a ver com *parras* e *uvas*. É especialmente nos casos mais idiomáticos que a tradução de provérbios se revela um problema particularmente interessante. No processo de tradução ou na busca de um equivalente de tradução, podem verificar-se várias situações: (i) uma tradução palavra-a-palavra existe, sobretudo no quadro da mesma comunidade linguístico-cultural alargada (e.g. o conjunto das línguas românicas); (ii) uma tradução parcial existe, em que um elemento ou mais elementos podem variar, mas o significado e as intenções comunicativas subjacentes se mantêm; (iii) é necessário recorrer a outro provérbio, sem qualquer relação do ponto de vista lexical. Embora muito interessante, não prosseguimos esta linha de trabalho, já abundantemente representada na literatura.

É desta perspetiva, que trata o provérbio como parte do léxico de uma língua, que discutimos a marginalidade e/ou centralidade do lugar do provérbio na gramática da língua: embora os gramáticos contemporâneos pareçam atribuir-lhe um estatuto quase anedótico, de curiosidade linguística, a realidade mostra que o emprego destas expressões atravessa vários géneros ou tipos textuais, sendo usado por todos os grupos demográficos da comunidade linguística. O seu número, pelo menos em português, não é despreciando. A quantidade de informação linguística (lexical, semântica, e pragmática) que se cristaliza numa expressão concisa e, por vezes, acutilante, como um provérbio, e o número de provérbios distintos, com significados e condições de uso diferentes, são tais que são de molde a fazer dos provérbios uma componente importante do léxico de muitas línguas, e, sem dúvida, do português.

¹<https://www.portalsplash.com/2010/04/dicionario-de-palavras-e-expressoes.html> [2020-02-28]

Vimos, também, como a perspetiva da gramática tradicional é particularmente omissa relativamente aos provérbios, mesmo que estes sejam usados como exemplos no tratamento de diferentes aspetos da gramática. Em contraponto, destacámos a contribuição do modelo do Léxico-Gramática que, pela sua conceção enquanto um modelo da gramática circunscrito às frases elementares da língua, aqui entendidas enquanto expressão de um predicado semântico, permite uma integração natural e, quanto a nós, adequada não só das frases livres mas também das expressões fixas/idiomáticas. Nesta perspetiva, a unidade de sentido e de análise gramatical não é palavra, mas sim a frase elementar, determinada pela construção de um elemento predicativo, e.g., um verbo, no caso de uma expressão com verbo pleno, distribucionalmente livre. O extenso trabalho já desenvolvido sobre o léxico-gramática das expressões fixas demonstrou, de forma abundante, que a descrição das frases fixas não difere substancialmente da das frases livres, ressalvadas as fortes restrições combinatórias verbo-argumento(s) que caracterizam as primeiras. Ora, os provérbios, na sua grande maioria, correspondem igualmente a frases em que, de facto, se verifica diferentes fenómenos de fixidez formal, mas que são, do ponto de vista estrutural/sintático, idênticas às construções livres. A estrutura sintática das frases proverbiais foi, por isso, um dos aspetos investigados nesta tese, tendo procurado elaborar uma taxonomia formal dos tipos de construções sintáticas encontradas, que dê, no entanto, conta da sua variação formal.

Aliás, a variação formal nos provérbios é um aspeto central nesta tese. Não só nos esforçámos por adotar meios formais de representação da variação que os provérbios podem apresentar, como procurámos desenvolver critérios linguisticamente motivados, que permitissem reunir sob o construto abstrato de *unidade paremiológica* o conjunto das variantes que é possível associar, com o mesmo significado e as mesmas condições pragmáticas de uso, às diferentes expressões. Assim, discutimos como, nas coletâneas de provérbios usuais, o modelo de representação formal dos provérbios, sob a forma de listagens, ainda que frequentemente introduza uma notação que representa a variação lexical ou a facultividade de um elemento, não dá conta, pelo de menos de uma forma completamente adequada, das unidades paremiológicas em que é possível agrupar diferentes variantes de um dado provérbio. Este desafio que os provérbios constituem para uma lexicografia mais tradicional é, no entanto, vencido pela adoção de outros métodos de representação, que aqui investigámos e que nos parecem mais eficientes, tanto mais que podem ter diferentes aplicações, nomeadamente no reconhecimento automático destas expressões em *corpora*.

Noutra dimensão, e na procura de um próprio das expressões proverbiais, bem como das complexas relações que estas estabelecem com as frases fixas/idiomáticas,

referimos, ainda que brevemente, alguns dos processos gramaticais que constroem o valor genérico provérbios, isto é, o seu tom sentencioso, de verdade intemporal. Tal prende-se com a identificação de provérbios que apresentam o que chamamos de *sujeito indefinido*, e em que, por meio de pronomes indefinidos que expressam o valor de grupos nominais genéricos, construídos com um pequeno conjunto de pronomes indefinidos, as fronteiras entre provérbio e frase livre tendem a se esbater.

Abordámos, ainda, as complexas relações que o provérbio pode estabelecer com o texto em que se insere, nomeadamente as funções discursivas/retóricas que exerce no discurso. Recuperámos o conceito de *citação* para considerar, como outros autores, que o provérbio é um texto *auto-contido*, isto é, um *texto dentro de outro texto*, sentido como *texto alheio*, e por isso mesmo, mantém em relação ao discurso uma certa autonomia (suspensão dos processos anafóricos pronominais, ou ausência de referência de grupos nominais definidos, por exemplo). Vimos, com Lopes (1992), como os provérbios podem ser usados na construção de um argumento ou como comentário, ou fazerem parte dos mecanismos de organização macro-textual, quando funcionam como sumário anafórico e/ou resumo catafórico do conteúdo do texto.

Finalmente, demos conta da utilização de provérbios, quer em aplicações terapêuticas e de diagnóstico de diversos tipos de afasia, quer no ensino de língua e cultura portuguesas, tanto na dimensão de língua materna como de língua estrangeira. Neste sentido, a delimitação de um “núcleo duro” dos provérbios da língua, o *mínimo paremiológico*, poderá motivar melhor o desenvolvimento desse tipo de ferramentas e aplicações, produzindo, esperamos, resultados mais adequados.

Contributos

Como dissemos acima, um dos principais objetivos desta tese consistiu em desenvolver métodos e ferramentas que permitissem identificar automaticamente provérbios em textos do português europeu. Este objetivo específico foi atingido, sendo hoje possível identificar automaticamente um grande número dos provérbios por nós coligidos, até quando estes não estão associados a quaisquer mecanismos introdutórios que muitas vezes os acompanham, e.g. *lá dizia a minha avó/o povo/o provérbio*, ou aspas.

Para este objetivo, foi essencial a constituição de uma base de dados, resultante da compilação de quatro coletâneas de provérbios, contendo mais de 114 mil expressões proverbiais. Cada provérbio/expressão encontra-se identificado de forma unívoca, o que permite determinar a sua fonte original. A base de dados poderá ser estendida, integrando outras coletâneas existentes. Esta listagem de provérbios serviu de base a praticamente todos os estudos que realizamos nesta investigação.

Utilizando o sistema UNITEX (Paumier 2016), baseado em léxicos e gramáticas locais para o processamento de *corpora* textuais, foram desenvolvidos diferentes conjuntos de grafos que tiveram por base:

- (i) Um conjunto de grafos formados apenas pelas palavras-chave de cada provérbio, mas aceitando um conjunto limitado de inserções (de 0 a 3 palavras e alguns sinais de pontuação) entre cada uma dessas palavra-chave. As palavras-chave de cada provérbio foram determinadas semiautomaticamente, após a remoção (também semiautomática) das *stopwords*. Estas são, sobretudo, palavras gramaticais, que foram removidas antes do processamento do texto. Neste caso, foram desenvolvidos 3 tipos de grafos diferentes: com 2, 3 e 4 palavras-chave. Nos casos em que os provérbios apresentavam mais do que 4 palavras-chave, utilizaram-se somente as primeiras quatro. Descartaram-se 124 provérbios da base de dados com menos de duas palavras-chave.
- (ii) Outro conjunto de grafos em que as variantes de cada provérbio se encontram reunidas sob uma mesma *unidade paremiológica* (UP), aceitando também algumas inserções entre certos constituintes. Estas variantes podem apresentar diferentes níveis de complexidade sintática e extensão lexical variável.
- (iii) Finalmente, um conjunto de grafos que representam as expressões introdutórias (formulísticas) e os termos associados ao conceito de *provérbio* (*aforismo*, *anexim*, *rifão*, etc.), com que muito frequentemente se introduzem os provérbios no discurso.

Dos resultados obtidos da aplicação destes grafos aos textos pôde concluir-se que:

- (a) Os grafos desenvolvidos a partir das palavras-chave dos provérbios têm a vantagem de poderem ser gerados semiautomaticamente e permitem capturar um grande número de expressões, até mesmo certas variações criativas de provérbios que muitas vezes se encontram nos textos, e.g. *A galinha da vizinha é **realmente** sempre melhor que a minha*. Por outro lado, pelo seu processo de construção, estes grafos apresentam a desvantagem de não aceitarem variação morfológica das palavras que constituem os provérbios e, pela sua natureza heurística, capturam muitas ocorrências espúrias, isto é, que não são efetivamente provérbios mas outras construções com as mesmas sequências de palavras-chave. Além disso, não delimitam adequadamente as fronteiras do provérbio, já que apenas capturam as sequências da primeira à última palavra-chave. Em síntese, a utilização destes grafos permite atingir uma boa cobertura, mas uma precisão inferior quando comparados com outros métodos.

- (b) Os grafos das variantes, por sua vez, permitem agrupar um grande número de provérbios e variantes sob uma mesma unidade paremiológica e capturam os provérbios em toda a sua extensão. Ainda que sejam grafos bastante precisos, têm como desvantagem não poderem ser gerados semiautomaticamente e só capturarem as expressões neles representadas. Como a elaboração destes grafos é feita manualmente, e dada a extensão do léxico a tratar ser considerável, ainda não foi possível atingir uma cobertura lexical inteiramente satisfatória.
- (c) Finalmente, os grafos que representam as expressões introdutórias e os termos sinónimos de *provérbio* têm a vantagem de aceitar variação morfossintática dos elementos das expressões introdutórias bem como dos termos pesquisados; mas apresentam a desvantagem de, nem sempre, as expressões introdutórias e os termos estarem associados a situações em que há, de facto, o uso de provérbios.

Foram também construídos automaticamente um conjunto de *scripts* com comandos **grep** para aplicação a textos simples, usando um sistema UNIX. Os comandos consistiram em expressões regulares para a procura de sequências de 2 a 5 palavras-chave, permitindo uma janela de 0 a 3 palavras entre cada palavra-chave e/ou uma lista de separadores autorizados (sinais de pontuação, aspas, etc.). Este método, que é semelhante ao acima descrito em (i), recuperou um grande número de falsos positivos, a maior parte resultante das pesquisas de sequências de 2 palavras-chave, devido a estas poderem muitas vezes corresponder as sequências que não constituem provérbios. Uma das limitações deste método resulta do facto de não permitir procurar as palavras-chave pelos seus lemas, e conseqüentemente, não permitir a captura de certas variações, sobretudo da flexão verbal.

Um outro objetivo a que nos tínhamos proposto respeita ao estabelecimento de uma tipologia formal sintática para os provérbios do português europeu. Neste sentido, foi desenvolvido um quadro de classificação formal (sintática), elaborado a partir da proposta inicialmente feita por Rassi et al. (2014) para uma coleção de cerca de +500 provérbios do português do Brasil, mas tendo, agora, uma maior base empírica (mais de 114 mil provérbios) e reformulando ou mesmo acrescentando alguns critérios formais. O objetivo desta classificação é o de reunir sob certas classes conjuntos de provérbios formalmente idênticos do ponto de vista sintático. Pretendeu-se, além disso, adotar um conjunto de critérios estritamente sintático, já que a classificação de Rassi et al. (2014) assentava, em primeiro lugar, no conceito demasiado abstrato de *proposição*, o que se revelou ser um critério de reprodutibilidade problemática, sobretudo por se tratar de um construto teórico e não de uma propriedade formal, diretamente observável. Por outro lado, o nosso objetivo era o de utilizar processos estritamente

mecânicos na aplicação do algoritmo de classificação e que pudessem ser reproduzidos por máquinas, a fim de classificar a extensa base de dados resultante da compilação de várias coletâneas de provérbios.

A classificação por nós proposta é assente nas seguintes classes formais de base (Tabela 10.1):

Tabela 10.1: *Classes formais dos provérbios do português europeu (Adaptado de Reis & Baptista 2019b)*

Classe	Construção	Exemplo
P1	Construções impessoais	<i>Depressa e bem não há quem</i>
P2	Construções predicativas	<i>Contas são contas</i>
P3	Construções sem verbo	<i>Pão pão, queijo queijo</i>
P4	Orações substantivas	<i>Quem tudo quer tudo perde</i>
P5	Orações subordinadas (adv. e adj.)	<i>Enquanto o pau vai e vem, folgam as costas</i>
P6	Orações comparativas	<i>Antes dobrar que quebrar</i>
P7	Orações coordenadas	<i>Cresce e aparece</i>
P8	Frases simples	<i>A verdade vem sempre ao de cima</i>

Algumas destas classes foram divididas de acordo com os constituintes internos dos provérbios. Estas subdivisões justificam-se essencialmente nas classes que apresentam um grande número de provérbios desse tipo. Disso é exemplo a classe P2, referente às construções predicativas, constituídas por frases simples que apresentam um verbo copulativo, que liga o sujeito a um outro constituinte. Esta classe é subdividida em função do elemento que é núcleo do predicativo do sujeito: um grupo nominal, um grupo adjetival, um grupo preposicional ou, ainda, um grupo adverbial. A classificação sintática proposta foi avaliada tendo por base a classificação levada a cabo, independentemente, por dois anotadores, que classificaram os provérbios do mínimo paremiológico, com base num conjunto inicial de diretivas de classificação/anotação. Foi observado um grau de concordância de 97,8% entre os anotadores. As diferenças de classificação observadas entre os anotadores permitiram melhorar as diretivas de anotação inicialmente estabelecidas. Concluímos, assim, que as diretivas de anotação elaboradas eram suficientemente claras, precisas e reprodutíveis. Desta forma, consideramos que estes critérios poderão ser utilizados para classificar automaticamente as restantes expressões proverbiais da base de dados (mais de 114 mil), ficando por desenvolver o conjunto de regras que, depois de implementadas no sistema de processamento STRING, deverão permitir classificar adequadamente todos (ou pelo menos a maioria) destes provérbios.

O estudo que desenvolvemos sobre as relações discursivas entre os provérbios e os textos em que estes se encontram permitiu-nos verificar que nem sempre é fácil determinar com que intenção comunicativa são os provérbios utilizados num dado discurso. Para podermos aferir a intenção comunicativa subjacente ao uso de provérbios desenvolvemos um inquérito por questionário que procurava elicitar o conhecimento que o falante tem sobre o contexto de uso do provérbios e, de modo indireto, determinar a sua intenção comunicativa. Foi estabelecida uma classificação de referência, elaborada independentemente por dois anotadores, que permitiu estabelecer uma comparação entre as respostas dos inquiridos e esse *golden standard*.

Ainda que os resultados não nos tenham permitido avançar de forma efetiva nessa matéria, permitiram-nos, no entanto, verificar que as pessoas têm mais facilidade em identificar o tópico a propósito do qual o provérbio é dito do que em atribuir um valor específico àquilo que o provérbio diz. Cerca de 70% dos respondentes do inquérito identificaram corretamente o tópico a propósito do qual o provérbio era invocado no discurso; enquanto somente cerca de 50% dos respondentes atribuíram corretamente um valor de polaridade (positivo-neutro-negativo) ao provérbio. O facto de as pessoas nem sempre conseguirem atribuir um valor de polaridade ao provérbio numa determinada situação comunicativa poderá ser um sinal de que, potencialmente, um mesmo provérbio poderá desempenhar diferentes funções pragmáticas/discursivas nos textos em que se encontram. Assim, a função específica que um provérbio desempenha dependerá do contexto discursivo em que está inserido.

Foi também em relação à polaridade atribuída aos provérbios que os 2 anotadores que serviram como referência para este estudo mais discordaram. Estes nem sempre conseguiram atribuir de forma consistente e reproduzível um valor de polaridade a cada provérbio, ou até a um mesmo provérbio num mesmo contexto de uso. A escala de polaridade poderá, eventualmente, não ser o melhor método para aferir a interpretação específica de alguns provérbios. O estudo das intenções comunicativas associadas ao uso de provérbios é pois um assunto que carece de um estudo mais aprofundado e as conclusões a que se chegou são apenas provisórias.

Um outro objetivo desta tese foi determinar o conjunto dos provérbios mais bem conhecidos e mais frequentemente utilizados pelos falantes do português europeu (o *mínimo paremiológico*). Para este fim, foi necessário adotar e combinar diferentes estratégias, de forma a garantir que os provérbios selecionados para integrar esse mínimo paremiológico são efetivamente aqueles que a grande maioria dos falantes dessa comunidade linguística conhece e usa.

Num primeiro momento, verificou-se quais os provérbios que surgiam mais vezes nas coletâneas que serviram de base a este estudo. Depois, dois anotadores,

independentemente, selecionaram os provérbios que consideravam mais usuais a partir do conjunto de todos os provérbios da base de dados. Após comparados os resultados desta seleção manual, definiu-se que os provérbios que tinham sido selecionados pelos dois anotadores seriam classificados de *nível 2*; os que tinham sido selecionados por apenas um dos anotadores seriam de *nível 1*; e os restantes provérbios, os não selecionados por nenhum dos anotadores, seriam de *nível 0*. A hipótese implícita a esta seleção era de que os provérbios selecionados pelos 2 anotadores seriam *muito usuais*, os selecionados por apenas um dos anotadores seriam apenas *usuais* e os que não foram selecionados seriam *não usuais*. Para confirmar esta hipótese foi desenvolvido e aplicado um inquérito por questionário.

Como não era exequível perguntar a uma parte representativa dos falantes do português europeu se conhecia cada um dos provérbios que estão atestados nas coletâneas de provérbios portugueses, selecionou-se aleatoriamente 100 provérbios (dos três níveis) para integrar um questionário. Nesse questionário, os respondentes indicaram *se conheciam e usavam o provérbio, se conheciam mas não usavam o provérbio*, ou se simplesmente *não conheciam o provérbio*. Cada uma destas respostas corresponderia, assim, aos níveis previamente definidos na referência construída manualmente.

Paralelamente a esta abordagem, verificou-se em diferentes tipos de *corpora* – manuais de português língua materna e de língua não materna, assim como *corpora* de textos jornalísticos – quais os provérbios que aí surgiam e qual a sua frequência, usando as ferramentas que desenvolvemos para identificar automaticamente provérbios em textos. Foi ainda verificada a frequência de uso dos provérbios em dois motores de pesquisa distintos – *Google* e *Bing*. A frequência dos provérbios nestes *browsers* apresenta uma forte correlação com a classificação manual, sendo os provérbios considerados como mais usuais os que, de facto, ocorrem com maior frequência na rede. Um outro questionário desenvolvido, que tinha como primeiro objetivo aferir a qualidade de um conjunto de jogos com provérbios, também nos permitiu validar a seleção feita pelos anotadores.

O resultado da aplicação destas diferentes estratégias permitiu-nos constituir o mínimo paremiológico (MP) do português europeu, que contém atualmente 318 unidades paremiológicas, abrangendo, no total, mais de 14.500 variantes. Estes provérbios são efetivamente as expressões mais frequentes e mais bem conhecidas pela generalidade dos falantes do português europeu. Algumas delas, ainda que não fossem propriamente provérbios (como, por exemplo, o bem conhecido slogan *Há mar e mar, há ir e voltar*), foram incluídas neste MP por serem utilizadas proverbialmente, com sentido genérico, e por apresentarem, como é o caso neste exemplo, diversas marcas formais distintivas do género proverbial (rima, paralelismo, aliteração, etc.). Além disso, em vários casos, a noção de autoria destas expressões já desapareceu, chegando

a surgir no discurso associadas às mesmas expressões formulaicas que muitas vezes introduzem os provérbios (*Lá diz o povo, Como dizia a minha avó, Bem diz o ditado, etc.*).

O mínimo paremiológico estabelecido é, sem dúvida, um dos subsídios que consideramos mais relevantes deste trabalho, na medida em que contribui para o melhor conhecimento das estruturas do léxico, e que poderá ter diversas aplicações. Considerando que os provérbios são utilizados como instrumento de diagnóstico/terapia de alguns distúrbios da linguagem resultantes de vários tipos de lesões; e que, muitas vezes, é pedido aos doentes afásicos que os interpretem, enunciem ou completem; é, pois, de grande importância que a seleção das expressões que são empregues nestas tarefas se faça a partir do conjunto daqueles que são comumente conhecidos.

Muitos dos provérbios selecionados para os testes de interpretação de provérbios não apresentam uma interpretação-padrão literal (cf. *Teste de rastreio DA de provérbios*, Santos 2009) e não são apresentados aos doentes no contexto discursivo onde ocorrem. É preciso reconhecer que, até para os indivíduos que não apresentam quaisquer afasias, poderá ser um desafio interpretar provérbios como *Pão, pão, queijo, queijo*, se não conhecerem previamente o provérbio e se não lhes for apresentado o contexto onde ocorre, que de algum modo lhes permita inferir/interpretar a situação que motivou o seu uso. Ao que nos parece, muitos dos testes com provérbios para sujeitos afásicos e que têm por objetivo medir a sua “capacidade de abstração (função executiva/linguagem-sintaxe)” são-lhes apresentados desprovidos de contexto. Repare-se no seguinte exercício retirado das instruções de aplicação do “Teste de rastreio de doença de Alzheimer com provérbios” (cf. *ibidem*: 80):

“Ler o provérbio em voz alta para o entrevistado e pedir que o interprete”

Ainda que os três provérbios associados a este exercício, v.g. *Quem semeia vento colhe tempestade, Cão que muito ladra não morde, Em terra de cego quem tem um olho é rei*, sejam bem conhecidos e usuais (pelo menos, para os falantes da variante do PE, como se verificou); não é de todo evidente que quem não conheça previamente estes provérbios os consiga interpretar se os mesmos não estiverem integrados num contexto discursivo adequado. Ora, uma das conclusões a que chegámos a partir do trabalho que desenvolvemos é que a função precisa de um provérbio (e logo assim a sua significação) parece depender mais do contexto do que do próprio provérbio em si. Posto isto, será necessário refletir não só na seleção dos provérbios a integrar nestes instrumentos de diagnóstico, como também na forma como poderão ser concebidas as tarefas que incluem provérbios. Esta será, pois, uma pista para trabalho futuro.

Até este momento, desenvolvemos algumas tipologias de jogos, assentes em processos linguísticos motivados, que poderão ser um primeiro passo no desenvolvimento de instrumentos de diagnóstico/terapia destas afasias. Alguns destes

jogos foram adaptados de forma a poderem vir a ser integrados na plataforma VITHEA-KIDS, em resposta à falta de aplicações específicas adaptadas às capacidades individuais, sintomas e desafios das crianças com autismo. Por razões que nos são alheias (sobretudo de ordem técnica), não foi ainda possível integrar na plataforma estes jogos, que apresentamos no Anexo 8.

Por outro lado, os jogos aqui desenvolvidos poderão ainda ser utilizados para o ensino-aprendizagem de português, quer enquanto língua materna, quer enquanto língua não materna. Acompanhando o movimento atual, em que a dimensão de auto-estudo ou estudo individual, remoto e assíncrono, se torna um paradigma de ensino e aprendizagem cada vez mais importante na sociedade da informação, a construção (semiautomática) deste tipo de recursos poderá contribuir para o desenvolvimento das atuais plataformas de ensino de língua assistido por computador.

Creemos que atingimos os objetivos inicialmente propostos, ainda que estejamos ciente de que muito que há a explorar e a fazer.

Epílogo

Os métodos de identificação automática de provérbios no seu estado atual de desenvolvimento, ainda que cumpram a sua função mais elementar, não dão conta (pelo menos de forma precisa) de muitas das situações de variação criativa da utilização de provérbios. Disso é exemplo o título que a seguir apresentamos retirado de um artigo recente publicado na edição *online* da revista *Maxima* (negritos nossos):

O sexo dos vizinhos é melhor do que o meu?

E se lhe dissermos que, afinal, a sua felicidade sexual não depende só do seu parceiro mas sim dos casais sexualmente felizes que a rodeiam? Descubra mais uma **variante do provérbio** em que se afirma que “a galinha da vizinha é melhor do que a minha”²

Se, para o leitor mais desatento, a referência implícita ao provérbio *A galinha da vizinha é sempre melhor do que a minha* não tivesse sido percebida, as linhas subsequentes do *lead* permitir-lhe-iam identificar esse provérbio, e, possivelmente, reconhecer, então, no título a referência ao provérbio. Ora, os métodos desenvolvidos neste estudo já permitem identificar a ocorrência do provérbio no *lead*, como se mostra na Figura 11.1, abaixo, retirada da saída do sistema STRING (Mamede et al. 2012)³:

²<https://www.maxima.pt/lifestyle/detalhe/o-sexo-dos-vizinhos-e-melhor-do-que-o-meu> [28-02-2020]

³<http://string.l2f.inesc-id.pt>

Referências Bibliográficas

- Abad, A., Pompili, A., Costa, A., Trancoso, I., Fonseca, J., Leal, G., . . . Martins, I. P. (2013). Automatic word naming recognition for an on-line aphasia treatment system. *Computer Speech and Language*, Elsevier, 27(6), pp. 1235-1248.
- Agno, F. (1960). Premessa a un repertorio di frasi proverbiali. In *Romance philology*, vol. 13, 3, pp. 242-264.
- Arif, M., & Abdullah, I. (2016). The impact of output communication on EFL learners' metaphor second language acquisition. *Social Sciences (Pakistan)* 11(9), pp. 1940-1947.
- 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.
- Baptista, J. (1994). *Estabelecimento e formalização de classes de nomes compostos*. (Dissertação mestrado não publicada). Universidade de Lisboa, Lisboa.
- Baptista, J. (2010). *Verba dicendi: A structure looking for verbs*. In T. Nakamura, É. Laporte, A. Dister, & C. Fairon (Eds.), *Les Tables. La grammaire du français par le menu. Mélanges en hommage à Christian Leclère*. Louvain-la-Neuve: CENTAL/Presses Universitaires de Louvain, pp. 11-20.
- Baptista, J. (2012). ViPEr: A Lexicon-Grammar of European Portuguese Verbs. In J. Radimsky (Ed.), *Proceedings of the 31st international conference on lexis and grammar*. Nové Hradý (Czech Republic), pp. 10-16: Università degli Studi di Salerno (Italy)/University of South Bohemia in Nové Hradý (Czech Republic).
- Baptista, J. (2013). ViPEr: uma base de dados de construções léxico-sintáticas de verbos do Português Europeu. In F. Silva, I. Falé, & I. Pereira (Eds.), *Actas do XXVIII Encontro da APL - Textos Seleccionados*. Lisboa: APL/Colibri, pp. 111-129.
- Baptista, J., Correia, A., & Fernandes, G. (2004). Frozen sentences in Portuguese: Formal Description for NLP. In *MWE '04 Proceedings of the Workshop on Multiword Expressions: Integrating Processing, Association for Computational Linguistics (ACL)*, Barcelona, Spain, pp. 72-79.
- Baptista, J., Fernandes, G., & Correia, A. (2005). Léxico-gramática das frases fixas do português europeu. Breve presentación. *Cadernos de Fraseoloxía Galega*, 7, Santiago de Compostela: Xunta de Galicia, pp. 41-53.

- Baptista, J., Fernandes, G., Talhadas, R., Dias, F., & Mamede, N. (2016). Implementing European Portuguese verbal idioms. In Gloria Corpas Pastor (Ed.) *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives*, Editions Tradulex, pp. 102-115.
- Baptista, J., & Mamede, N. (2020). *Dicionário Gramatical de Verbos do Português Europeu*. Universidade do Algarve.
- Baptista, J., Mamede, N., & Gomes, F. (2010). Auxiliary Verbs and Verbal Chains in European Portuguese. In T. Pardo, A. Branco, A. Klautau, R. Vieira, V. De Lima (Eds.) *Computational Processing of the Portuguese Language. PROPOR 2010. Lecture Notes in Computer Science*, vol 6001. Springer, Berlin, Heidelberg, pp. 110-119.
- Baptista, J., Marques, J., Fernandes, G., Guerra, J., Mariano, A., Marques, J., ... Reis, S. (2018). Emprego dos pronomes clíticos do português europeu em corpora de aprendizagem de português língua estrangeira. VII SIELP – Simpósio Internacional de Ensino da Língua Portuguesa, U. Minho, Braga, 16-18 julho 2018 (2018, aceite para publicação).
- Baptista, J. A., Santiago, A., Almeida, D., Antunes, P., & Gaspar, R. (2011). *Programa de português L2 para alunos surdos – Ensinos Básico e Secundário*. Ministério da Educação.
- Basso, A. (1992). Prognostic factors in aphasia. *Aphasiology*, 6(4), pp. 337-348.
- Bechara, E. (2010). *Gramática escolar da língua portuguesa*. 2ª edição (edição ampliada e atualizada pelo novo Acordo Ortográfico), Editora Nova Fronteira.
- Benton, A. (1968). Differential behavioral effects in frontal lobe disease. *Neuropsychologia* 6(1), pp. 53-60.
- Benveniste, E. (1962). Les niveaux de l'analyse linguistique. In PLG, I, (1966), pp. 119-131.
- Bhogal, S., Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke* 34(4), pp. 98-993.
- Bick, E. (2000). *The parsing system "PALAVRAS": Automatic grammatical analysis of Portuguese in a constraint grammar framework*. Aarhus, Denmark: Dr.phil. Thesis. Aarhus University. Aarhus, Denmark: Aarhus University Press.
- Blanco, X. (2010). Los frasemas composicionales pragmáticos. In S. Mejri y P. Mogorrón (Eds.). *Opacité, Idiomaticité, Traduction*. Alacant: Universitat d'Alacant, pp. 19-32.
- Bruckschen, M., Muniz, F., Souza, J., Fuchs, J., Infante, K., Muniz, M., ... Aluísio, S. (2008). Anotação linguística em XML do corpus PLN-BR. São Carlos (SP), NILC-ICMC-USP, *Série de Relatórios do NILC*.
- Buescu, H., Morais, J., Rocha, M., & Magalhães, V. (2015). *Programa e Metas Curriculares de Português no Ensino Básico*. Governo de Portugal – Ministério da Educação e da Ciência.

- Cardoso, P., Maziero, E., Jorge, M., Seno, E., Di Felippo, A., Rino, L., ... Pardo, T. (2011). CSTnews-a discourse-annotated corpus for single and multi-document summarization of news texts in Brazilian Portuguese. *Proceedings of the 3rd RST Brazilian Meeting*, pp. 88-105.
- Carneiro, A. (2016). *Descrição e classificação das expressões cristalizadas com ser e estar do português do Brasil*. (Dissertação mestrado não publicada). Universidade Federal de São Carlos, São Carlos.
- Carvalho, P. (2007). *Análise e representação de construções adjetivais para processamento automático de texto. Adjetivos intransitivos humanos*. (Tese doutoramento não publicada). Universidade de Lisboa, Lisboa.
- Cazelato, S. (2003). *A interpretação de provérbios equivalentes por afásicos: um estudo enunciativo*. (Dissertação mestrado não publicada). Campinas, SP.
- Cazelato, S. (2008). *A interpretação de provérbios parodiados por afásicos e não afásicos*. (Tese doutoramento não publicada). Universidade Estadual de Campinas, Instituto de Estudos da Linguagem, Campinas, SP.
- Chacoto, L. (1994). *Estudo e formalização das propriedades léxico-sintáticas das expressões fixas proverbiais*. (Dissertação mestrado não publicada). Universidade de Lisboa, Lisboa.
- Chacoto, L. (2007). A sintaxe dos provérbios. As estruturas quem/quien en português e espanhol. *Cadernos de Fraseoloxía Galega, 9, Santiago de Compostela: Centro Ramón Piñero – Xunta de Galicia*, pp. 31-53.
- Chacoto, L. (2008). Vale mais um gosto na vida que três vinténs na algibeira – Las estructuras comparativas en los proverbios portugueses. In Conde Tarrío, Germán (Ed.). *Aspectos formales y discursivos de las expresiones fijas*, Frankfurt am Main u.a.: Peter Lang, pp. 87-103.
- Chacoto, L. (2010). Não há rifão velho, se é dito a propósito – La condición en los refranes portugueses. In Korhonen, J. Mieder, W. Piirainen, E. Piñel, R. (Eds.), *Atas do Congresso Internacional Europhras 2008*, Finlândia: Universidade de Helsínquia, pp. 58-66.
- Chaika, E. (2000). *Linguistics, pragmatics and psychotherapy – A guide for therapists*. London and Philadelphia: Whurr Publishers.
- Charteris-Black, J. (1995). Proverbs in communication. *Journal of Multilingual and Multicultural Development 16*(4), pp. 259-268.
- Clark, A., Fox, C., & Lappin, S. (Eds.) (2010). *The Handbook of Computational Linguistics and Natural Language Processing*. Wiley-Blackwell.
- Conenna, M. (1988). *Sur un lexique-grammaire comparé de proverbes*. *Langages, (90)*, pp. 99-116.
- Conselho da Europa. (2001). *Quadro europeu comum de referência para as línguas – aprendizagem, ensino, avaliação*. Porto: Edições ASA.

- Correia, R. (2010). *Automatic question generation for REAP.PT Tutoring System*. (Dissertação mestrado não publicada). Universidade Técnica de Lisboa, Instituto Superior Técnico, Lisboa.
- Correia, R., Baptista, J., Eskenazi, M., & Mamede, N. (2012). Automatic generation of cloze question stems. In Caseli H., Villavicencio A., Teixeira A., Perdigão F. (Eds.), *Computational Processing of the Portuguese Language. PROPOR 2012. Lecture Notes in Computer Science*, vol 7243. Springer, Berlin, Heidelberg, pp. 168-178.
- Correia, R., Baptista, J., Mamede, N., Trancoso, I., & Eskenazi, M. (2010). Automatic generation of cloze question distractors. In *Proceedings of the Workshop on Second Language Studies: Acquisition, learning, education and technology*. Tokyo, Japan.
- Costa, J. (1999). *O Livro dos Provérbios Portugueses*. Lisboa: Editorial Presença.
- Costa, J. (2008). *O Advérbio em Português Europeu*. Lisboa: Colibri.
- Council of Europe. (2001). *Common european framework of reference for languages: Learning, teaching, assessment*. Language Policy Unit, Strasbourg, www.coe.int/lang-CEFR.
- Coxhead, A., Nation, P., & Sim, D. (2015). Measuring the vocabulary size of native speakers of English in New Zealand Secondary schools. *Journal of Educational Studies* 50, pp. 121-135.
- Cunha, C., & Cintra, L. (1984). *Nova Gramática do Português Contemporâneo*. (5^a ed. 1999) Lisboa: Edições Sá da Costa.
- Curto, P. (2014). *Classificador de textos para o ensino de português como segunda língua*. (Dissertação mestrado não publicada). Universidade de Lisboa / Instituto Superior Técnico, Lisboa.
- Dias, M. (2009). *A articulação hipotática em construções proverbiais justapostas*. (Dissertação mestrado não publicada). Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- Dias-da-Silva, B. (2013). Modelagem linguístico-computacional de léxicos. In: Laporte, Éric; Smarsaro, Aucione; Vale, Oto A. (Orgs.) *Dialogar é preciso: linguística para processamento de línguas*. Vitória: PPGEL/UFES, pp. 89-103.
- Direção de Serviços de Língua e Cultura. (2017). *Referencial Camões PLE*. Lisboa: Camões, Instituto da Cooperação e da Língua I.P.
- Duarte, T. (2013). *Cinco unidades didáticas com provérbios: a produtividade dos provérbios na síntese e memorização de conhecimentos*. Relatório Científico, Porto: Universidade do Porto.
- Đurčo, P. (2004). Slovak proverbial minimum: The empirical evidence. In Cs. Földes (Ed.), *Res humanae proverborum et sententiarum*. Ad honorem Wolfgangi Mieder, Tübingen: Gunter Narr, pp. 59-69.

- Ďurčo, P. (2005a). Empirisch- und Korpusbasierte Untersuchungen der Sprichwörter [Estudos empíricos e baseados em corpus de provérbios]. *Zeitschrift für germanistische Sprach- und Literaturwissenschaft in der Slowakei* 3, pp. 47-57.
- Ďurčo, P. (2005b). Paremiologické minimum slovenčiny. Výsledky a porovnania. [Mínimo Paremiológico do eslovaco. Resultados e comparações] In R. Blatná & V. Petkevič (Eds.), *Jazyky a jazykověda. Sborník k 65. narozeninám prof. PhDr. Františka Čermáka, DrSc.* Praha: Ústav Českého národního korpusu, pp. 45-61.
- Ďurčo, P. (2006). *Methoden der Sprichwortanalysen oder Auf dem Weg zum Sprichwörter-Optimum.* [Métodos de análise do provérbio ou no caminho para o ótimo de provérbio] In A. Häcki Buhofer and H. Burger (Eds.), *Phraseology in Motion I. Methoden und Kritik. Akten der Internationalen Tagung zur Phraseologie* (Basel, 2004). (Phraseologie und Parömiologie 19), pp. 3-20.
- Ďurčo, P. (2007). Paremiologické optimum slovenčiny. [Ótimo paremiológico do eslovaco] In W. Chlebda (Ed.), *Frazeologia a językowe obrazy świata przelomu wieków*, Opole: Uniwersytet Opolski, pp. 171-177.
- Ďurčo, P. (2015). Empirical research and paremiological minimum. In Hrisztova-Gotthardt, H. and Aleksa Varga, M. (Eds.) *Introduction to Paremiology. A Comprehensive Guide to Proverb Studies.* De Gruyter, pp. 183-205.
- Ďurčo, P., & Meterc, M. (2014). Empirične paremiološke raziskave tipov ekvivalentnosti in suprasemantičnih razlik v slovenščini in slovaščini [Investigações empíricas paremiológicas sobre os tipos de equivalência e diferenças supra-semânticas em esloveno e eslovaco]. *Slavia centralis*, pp. 20-36.
- Faria, I. (2003). O uso da língua, interação verbal e texto. In Mateus et al. *Gramática da Língua Portuguesa.* (5ª edição revista e aumentada). Lisboa: Caminho, pp. 55-84.
- Fellbaum, C. (Ed.). (1998). *Wordnet: An electronic lexical database.* Cambridge MA: MIT Press.
- Fernandes, G. (2007). *Léxico-gramática das frases fixas do português europeu – construções intransitivas.* (Dissertação mestrado não publicada). Universidade do Algarve, Faro.
- Ferreira, A. (2011). *O provérbio no ensino aprendizagem do PLE. Um pequeno passo para os alunos, um grande passo para uma competência comunicativa e intercultural.* (Dissertação mestrado não publicada). Universidade do Porto, Porto.
- Ferreira, H., & Vieira, M. (2013). O trabalho com o gênero provérbio em sala de aula. *Revista Línguas & Letras*, 14(26), pp. 1-18.
- Figueiredo, G. (2012). *O gênero proverbial na imprensa: usos e funções retóricas.* (Tese doutoramento não publicada). Universidade Federal de Pernambuco, Recife.
- Galvão, A., Baptista, J., & Mamede, N. (2019a). New developments on processing European Portuguese verbal idioms. *12th Symposium in Information and Human Language Technology*, October, 15-18, Salvador, BA (Brazil), pp. 229-238.

- Galvão, A., Baptista, J., & Mamede, N. (2019b). Processing European Portuguese Verbal Idioms: From the Lexicon-Grammar to a Rule-based Parser. In G. Corpas Pastor, R. Mitkov, M. Kunilovskaya, L. León, M. Araceli (Eds.) *Computational and Corpus-based Phraseology. Proceedings of the Third International Conference EUROPHRAS 2019*, September, 25-27, Malaga (Spain), Tradulex, pp. 70-77.
- Gangur, M. (2011). Automatic generation of cloze questions. *Proceedings of the 3rd International Conference on Computer Supported Education*, Volume 1, Noordwijkerhout, Netherlands, 6-8 May.
- García-Page Sánchez, M. (2008). *Introducción a la fraseología española. Estudio de las locuciones*. Barcelona: Anthropos.
- Gasanova, M., & Taibova, L. (2016). Revisiting paremiological units of the Tabasaran language. *Oriental Studies*, 25(3), 99-105. Doi: 10.22162/2075-7794-2016-25-3-99-105.
- Gibbs, R., & Beitel, D. (1995). What proverb understanding reveals about how people think. *Psychol Bull.* 118(1), pp. 133-154.
- González Rey, I. (2002). *La phraséologie du français*. Toulouse: Presses Universitaires du Mirail.
- Gorham, D. (1956). A proverb test for clinical and experimental use. *Psychological Reports*, 2(Mono Suppl 1) pp. 1-12.
- Gross, G. (1988). Degré de figement des noms composés. *Langages* 90, Larousse, Paris, pp. 57-72.
- Gross, G. (1996). Prédicats nominaux et compatibilité aspectuelle. *Langages*, 30^e année, 121, Les supports, pp. 54-72.
- Gross, M. (1975). *Méthodes en syntaxe. Régime des constructions complétives*. Paris: Hermann.
- Gross, M. (1982). Une classification des phrases figées du français. *Revue Québécoise de Linguistique*, 11(2), pp. 151-185.
- Gross, M. (1988a). *Les limites de la phrase figée*. *Langages*, 23^e année, 90. Les expressions figées, pp. 7-22.
- Gross, M. (1988b). Methods and tactics in the construction of a lexicon-grammar. *Linguistics in the morning calm 2*. Select papers from SICOL 1986, Hanshin Publishing Company, Seoul, pp. 177-197.
- Gross, M. (1994). The Lexicon Grammar of a Language: Application to French. (R.E.A)sher. *Encyclopedia of Language and Linguistics*, Pergamon, halshs-00278309, pp. 2195-2205.
- Gross, M. (1996). *Grammaire Transformationnelle du Français. Syntaxe de l'adverbe*. Paris: ASSTRIL.

- Grosso, M. (2011). *QuaREPE Quadro de Referência para o Ensino Português no Estrangeiro – Documento Orientador*. Disponível em https://www.dge.mec.pt/sites/default/files/EEstrangeiro/2012_quarepe_docorientador.pdf.
- Harris, Z. (1964). The Elementary Transformations. In Hiz H. (Ed.) 1981 *Papers on Syntax*. Synthese Language Library (Text and Studies in Linguistics and Philosophy), vol 14. Springer, Dordrecht, pp. 211-235.
- Harris, Z. (1965). Transformational Theory. In Hiz H. 1981 (Eds.) *Papers on Syntax*. Synthese Language Library (Text and Studies in Linguistics and Philosophy), vol 14. Springer, Dordrecht, pp. 236-280.
- Harris, Z. (1982). *A Grammar of English on Mathematical Principles*. New York: John Wiley & Sons.
- Harris, Z. (1991). *A Theory of Language and Information – A Mathematical Approach*. Oxford: Clarendon Press.
- Hoshino, A., & Nakagawa, H. (2005). A real-time multiple-choice question generation for language testing: A preliminary study. In *Proceedings of the second workshop on building educational applications using NLP, Association for Computational Linguistics*, pp. 17-20.
- HRISZTOVA-GOTTHARDT, H., & VARGA, M. (Eds.). (2015). *Introduction to Paremiology: A Comprehensive Guide to Proverb Studies*. Berlin: DeGruyter.
- Iriarte Sanromán, Á. (2001). *A unidade lexicográfica: palavras, colocações, frases, pragmatemas*. Braga: Centro de Estudos Humanísticos – Universidade do Minho.
- Jakobson, R. (1987). *Language in Literature*. Harvard University Press.
- Kent, R. (Ed.). (2004). *The MIT Encyclopedia of Communication Disorders*. MIT Press.
- Lamiroy, B. (1998). Le Lexique-Grammaire. Essai de synthèse. In B. Lamiroy (Ed.) *Travaux de Linguistique: Le Lexique-Grammaire* (Vol. 37). Duculot.
- Léger, A., Démonet, J., Ruff, S., Aithamon, B., Touyeras, B., Puel, M., ... Cardebat, D. (2002). Neural substrates of spoken language rehabilitation in an aphasic patient: An fMRI study. *NeuroImage*, 17, pp. 174-183.
- Leiria, I. (Coord.). (2008). *Orientações Programáticas de Português Língua não Materna (PLNM) – Ensino Secundário*. Ministério da Educação.
- Lopes, A. (1992). *Texto proverbial português – Elementos para uma análise semântica e pragmática*. (Tese doutoramento não publicada) Faculdade de Letras da Universidade de Coimbra, Coimbra.
- Lopes, A. (1998). Funções discursivas do provérbio em textos de imprensa. *Estudos de Literatura Oral, ELO*, IV, pp. 115-126.

- Machado, J. (2011). *O Grande Livro dos Provérbios*. (4ª ed.), Alfragide: Casa das Letras.
- Mamede, N., Baptista, J., Diniz, C., & Cabarrão, V. (2012). STRING: A Hybrid Statistical and Rule-Based Natural Language Processing Chain for Portuguese. Caseli, H.; Villavicencio, A.; Teixeira, A. and Perdigão, F. (Eds.), *Computational Processing of the Portuguese Language, Proceedings of the 10th International Conference*, PROPOR 2012 Demo Sessions, vol. Demo Session.
- Mann, W., & Thompson, S. (1983). Relational proposition in discourse. *ISI: Information Sciences Institute of University of Southern California*, Los Angeles, ISI/RR, pp. 83-115.
- Mann, W., & Thompson, S. (1988). Rhetorical structure theory. Toward a functional theory of text organization. *Text*, 8(3), pp. 243-281.
- Marques, J. (2013). *Anaphora resolution in Portuguese – An hybrid approach* (Unpublished master's thesis). Instituto Superior Técnico, Lisboa.
- Marrafa, P. (2001). *Wordnet do português: uma base de dados de conhecimento linguístico*. Lisboa: Instituto Camões.
- Martins, E. (1997). *Manual de redação e estilo*. (3ª ed.), O Estado de S. Paulo, São Paulo.
- Martins, P. (2010). Do provérbio em contexto didático: proposta de trabalho. *Paremia*, 19, pp. 93-102.
- Mateus, M., Brito, A., Duarte, I., Faria, I., Frota, S., Matos, G., . . . Villalva, A. (2003). *Gramática da língua portuguesa* (5ª ed. revista e aumentada). Lisboa: Caminho.
- Maziero, E. (2016). *Análise retórica com base em grande quantidade de dados*. (Tese doutoramento não publicada). Universidade de São Paulo, São Paulo.
- Maziero, E., Pardo, T., Di Felippo, A., & Dias-da Silva, B. (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), pp. 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), pp. 387-396.
- Meira, A. (2011). *A articulação de orações em provérbios do português em uso: uma análise das relações retóricas*. (Dissertação mestrado não publicada). Universidade Federal de Minas Gerais, Belo Horizonte.
- Meira, A. (2013). Amigos, amigos, negócios à parte: *uma análise da estrutura retórica*. *Anais do SILEL*. 3(1). Uberlândia: EDUFU, pp. 1-8.
- Meira, A. (2015). Casa de ferreiro, espeto de pau: *uma análise das relações retóricas a partir do uso dos provérbios como estratégia argumentativa em textos da internet*. (Tese doutoramento não publicada). Universidade Federal de Minas Gerais, Belo Horizonte.

- Mel'čuk, I. (1995). Phrasemes in Language and Phraseology in Linguistics. In M. Everaert, E. J. van der Linden, A. Schenk, and R. Schreuder (Eds.), *Idioms: Structural and Psychological Perspectives*. Hillsdale, New Jersey, Hove, UK: Lawrence Erlbaum Associates, pp. 167-232.
- Mel'čuk, I., & Polguère, A. (2007). *Lexique actif du français – l'apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français*. Bruxelles: De Boeck.
- Mendes, A. (2013). Organização textual e articulação de orações. In Paiva Raposo et al. (Org.). *Gramática do Português*. Vol. II. Lisboa: Calouste Gulbenkian, pp. 1691-1755.
- Mendonça, V., Coheur, L., & Sardinha, A. (2015). VITHEA-Kids: a Platform for Improving Language Skills of Children with Autism Spectrum Disorder. *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility, ASSETS '15*, New York, NY, USA, pp. 345-346.
- Meterc, M. (2012). Online questionnaire providing information on most well-known and well-understood proverbs in Slovene language. In V. Jesenšek et al (Ed.), EUROPHRAS 2012 Maribor. *Phraseologie und Kultur*, Maribor, pp. 195-205.
- Meterc, M. (2014). *Primerjava paremiologije v slovenskem in slovaškem jeziku na osnovi paremiološkega optimuma*. [Comparação da paremiologia em esloveno e eslovaco com base no parâmetro ótimo paremiológico] (Unpublished doctoral dissertation). Filozofska fakulteta Univerze v Ljubljani.
- Mieder, W. (1992). Paremiological minimum and cultural literacy. In Simon J. Bronner (Ed.). *Creativity and Tradition in Folklore: New Directions*. Logan, Utah: Utah State University Press, pp. 185-203.
- Mieder, W. (1994). *Wise Words (RLE Folklore)*. London: Routledge.
- Mieder, W. (2004). *Proverbs – A Handbook*. London: Greenwood Press.
- Mitkov, R. (2002). *Anaphora resolution*. Longman, Studies in Language and Linguistics.
- Molinier, C., & Lévrier, F. (2000). *Grammaire des adverbes: Description des formes en -ment*. Genève: Droz.
- Morato, E. (2010). *A semiologia das afasias: perspectivas linguísticas*. São Paulo: Cortez.
- Moreira, A. (1996). *Provérbios Portugueses*. Lisboa: Editorial Notícias.
- Musso, M., Weiller, C., Kiebel, S., Müller, S., Bülow, P., & Rijntjes, M. (1999). Training-induced brain plasticity in aphasia. *Brain*, 122, pp. 1781-1790.
- Neves, O. (2000). *Dicionário de Expressões Correntes*. Lisboa: (2ª ed.) Editorial Notícias.

- Nóbrega, F., & Pardo, T. (2017). Update summarization for Portuguese. In *Proceedings of the 6th Brazilian Conference on Intelligent Systems (BRACIS)*, IEEE, Uberlândia, pp. 348-353.
- Nóbrega, V., & Miyagawa, S. (2015). The precedence of syntax in the rapid emergence of human language in evolution as defined by the integration hypothesis. *Frontiers in Psychology*, 6, Article 271.
- Norrick, N. (1985). *How proverbs mean. Semantic studies in English proverbs*. Berlin: Mouton Publishers.
- Oliveira, H., & Gomes, P. (2014). ECO and Onto.PT: A flexible approach for creating a Portuguese wordnet automatically. In *Language Resources and Evaluation Springer*, 48(2), pp. 373-393.
- Oliveira, H., Santos, D., Gomes, P., & Seco, N. (2008). PAPEL: a dictionary-based lexical ontology for Portuguese (Vol. 5190). António Teixeira, Vera Lúcia Strube de Lima, Luís Caldas de Oliveira and Paulo Quaresma (Eds.), *Computational Processing of the Portuguese Language*, 8th International Conference, Proceedings (PROPOR 2008), (Aveiro, Portugal, 8-10 de Setembro de 2008), Springer Verlag.
- Palma, C. (2009). *Expressões fixas adverbiais: Descrição léxico-sintáctica e subsídios para um estudo contrastivo português-espanhol*. (Dissertação mestrado não publicada). Universidade do Algarve, Faro.
- Pardo, T. (2005). *Métodos para Análise Discursiva Automática*. (Tese doutoramento não publicada). Universidade de São Paulo, São Carlos-SP.
- Pardo, T., Nunes, M., & Rino, L. (2004). DiZer: An Automatic Discourse Analyzer for Brazilian Portuguese. *Proceedings of the 17th Brazilian Symposium on Artificial Intelligence – SBIA (Lecture Notes in Artificial Intelligence 3171)*, pp. 224-234.
- Parente, S. (2005). *O Livro dos Provérbios*. Lisboa: Editora Âncora.
- Parsian, M. (2015). *Data algorithms: Recipes for scaling up with Hadoop and Spark*. O'Reilly Media, Inc.
- Paumier, S. (2003). *De la Reconnaissance de Formes Linguistiques a l'Analyse Syntaxique*. (Unpublished doctoral dissertation). IGM – Institut Gaspard-Monge, Université de Marne-la-Vallée, Paris.
- Paumier, S. (2014). *Unitex 3.1 – User manual*. Université de Paris-Est/Marne-la-Vallée – Institut Gaspard Monge, Noisy-Champs.
- Paumier, S. (2016). *Unitex 3.1 – User manual*. Université de Paris-Est/Marne-la-Vallée – Institut Gaspard Monge, Noisy-Champs.
- Peck, K., Moore, A., Crosson, B., Gaiefsky, M., Gopinath, K., White, K., & Briggs, R. (2004). Functional magnetic resonance imaging before and after aphasia therapy: shifts in hemodynamic time to peak during an overt language task. *Stroke*, 35(2), pp. 554-559.

- Pellegrini, T., Ling, W., Silva, A., Trancoso, I., Correia, R., Baptista, J., & 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 2012)*, pp. 538-543.
- Pereira, A. (2015). *Competência linguística em português europeu língua materna e língua não materna: Aquisição, ensino e aprendizagem de expressões idiomáticas*. (Tese doutoramento não publicada). Universidade do Minho, Minho.
- Perkins, D. (1999). *The many faces of constructivism*. *Educational Leadership*, 5(3), pp. 6-11.
- Permjakov, G. (1973). On the paremiological level and premiological minimum of language. *Proverbium*, 22, pp. 862-863.
- Pianta, E., Bentivogli, L., & Girardi, C. (2002). MultiWordNet: developing an aligned multilingual database. *Proceedings of the 1st International WordNet Conference*, January 21-25, 2002, Mysore, India, pp. 293-302.
- Pies, R. (1994). *Clinical manual of psychiatric diagnosis and treatment: A biopsychosocial approach*. Washington DC: American Psychiatric Press.
- Pinheiro, M. (2015). *Por um dicionário eletrônico de pragmatemas do português brasileiro: levantamento, descrição e categorização*. (Tese doutoramento não publicada) Universidade Federal do Ceará, Fortaleza, CE.
- Ranchhod, E. (1991). Frozen adverbs – Comparative forms 'Como C' in Portuguese. *Linguisticae Investigationes*, XV(1), 141-170.
- Ranchhod, E. (2003). O Lugar das Expressões 'Fixas' na Gramática do Português, In Ivo Castro & Inês Duarte (Orgs.). *Razões e Emoção*, II, Lisboa, pp. 239-254.
- Ranchhod, E., Mota, C., & Baptista, J. (1999). A computational lexicon of Portuguese for automatic text parsing. *Proceedings of SIGLEX99: Standardizing Lexical Resources*, 37th Annual Meeting of the ACL College Park, Maryland, USA, pp. 74-80.
- Raposo, E. (2013). Estrutura da frase. In Paiva Raposo et al. (Org.). *Gramática do Português*. Vol. I. Lisboa: Calouste Gulbenkian, pp. 313-398.
- Rassi, A., Baptista, J., & Vale, O. (2014). Automatic Detection of Proverbs and their Variants. In *Proceedings of the Symposium on Languages, Applications and Technologies (SLATE'14)*. Leibniz (Germany): Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, pp. 235-249.
- Rassi, A., Santos-Turati, C., Baptista, J., Mamede, N., & Vale, O. (2014). The fuzzy boundaries of operator verb and support verb constructions with *dar* "give" and *ter* "have" in Brazilian Portuguese. *Workshop on Lexical and Grammatical Resources for Language Processing (LG-LP 2014)*, COLING 2014, Dublin, Ireland, pp. 92-101.
- Reis, S. (2014). *A correspondência entre provérbios e expressões fixas no português europeu*. (Dissertação mestrado não publicada). Universidade do Algarve, Faro.

- Reis, S., & Baptista, J. (2016a). 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, pp. 427-446.
- Reis, S., & Baptista, J. (2016b). O uso de provérbios no ensino de português. In Soares, R. & Lauhakangas, O. (Eds.) *10th Interdisciplinary Colloquium on Proverbs*, Actas ICP16 Proceedings. Tavira: AIP-IAP, 2017, pp. 521-538.
- Reis, S., & Baptista, J. (2016c). Portuguese Proverbs: Types and Variants. In Gloria Corpas Pastor (Ed). *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives*. Geneva: Editions Tradulex, pp. 208-217.
- Reis, S., & Baptista, J. (2017a). Estimating lexical availability of European Portuguese proverbs. In Mitkov R. (Ed.) *Computational and Corpus-Based Phraseology. EUROPHRAS 2017. Lecture Notes in Computer Science*, vol 10596. Springer, Cham, pp. 232-244.
- Reis, S., & Baptista, J. (2017b). Os provérbios em manuais de ensino de português língua não materna. In Vlândia Pinheiro & Gustavo Henrique Paetzold (Eds.) *Proceedings of Symposium in Information and Human Language Technology* Uberlandia, MG, Brazil, October 2-5, 2017, Sociedade Brasileira de Computação, pp. 247-255.
- Reis, S., & Baptista, J. (2018). Estudo das estruturas proverbiais com sujeito indefinido – provérbio ou *EF*? In Soares, R. & Lauhakangas, O. (Eds.) *12th Interdisciplinary Colloquium on Proverbs*, Atas ICP18 Proceedings, Tavira (aceite para publicação).
- Reis, S., & Baptista, J. (2019a). As intenções comunicativas no uso de provérbios. In Soares, R. & Lauhakangas, O. (Eds.) *13th Interdisciplinary Colloquium on Proverbs*, Atas ICP19 Proceedings, Tavira (aceite para publicação).
- Reis, S., & Baptista, J. (2019b). Descrição e classificação sintática das expressões proverbiais do português europeu. *Diacrítica* 33(2), pp. 324-334.
- Reis, S., & Baptista, J. (2020). Determinação de um mínimo paremiológico do português europeu. *Acta Scientiarum. Language and Culture*, 42(2), e52114. <https://doi.org/10.4025/actascilangcult.v42i2.52114>.
- Reis, S., Baptista, J., & Mamede, N. (2019). *Automatic Syntactic Classification of European Portuguese Proverbs*. Comunicação apresentada em International Conference 'Computational and Corpus-based Phraseology' EUROPHRAS 2019, Málaga, 25-27 setembro 2019 (2019, aceite para publicação).
- Reis, S., Baptista, J., & Mamede, N. (2020). Processing Proverb Variation in (European) Portuguese: Integrating the Paremiological Minimum in STRING. Comunicação apresentada em *Por Outras Palavras POP2 @ PROPOR 2020*, In *Other Words, 2nd Workshop on Linguistic Tools and Resources for Paraphrasing in Portuguese*, collocated with International Conference on the Computational Processing of Portuguese (PROPOR 2020), Colégio do Espírito Santo, Évora, (Portugal), March 2-4, 2020. (não publicada).

- Reis, S., Pompili, A., Abad, A., & 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).
- Rezaei, A. (2012). Rhetorical function of proverbs based on literary genre. *The Procedia – Social and Behavioral Sciences* 47, Elsevier Ltd, pp. 1103-1108.
- Rocha, P., & Santos, D. (2000). “CETEMPúblico: Um *corpus* de grandes dimensões de linguagem jornalística portuguesa”. In M. G. Volpe Nunes (Ed.), *Actas do V Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR’2000)*, (Atibaia, São Paulo, Brasil, 19 a 22 de Novembro de 2000), pp. 131-140.
- Salbego, N., & Osborne, D. (2016). Schema activation through pre-reading activities: teaching proverbs in L2. *BELT – Brazilian English Language Teaching Journal* 7(2), pp. 175-188.
- Santos, D., & Rocha, P. (2001). Evaluating CETEMPúblico, a free resource for Portuguese. *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics* (Toulouse, 9-11 de julho de 2001), pp. 442-449.
- Santos, M. (2009). *Elaboração e validade do teste de rastreio de doença de Alzheimer com provérbios (TRDAP) para indivíduos a partir de sessenta anos*. (Tese doutoramento não publicada). Universidade Federal do Rio Grande do Norte, Natal.
- Santos, M., Sougey, E., & Alchieri, J. (2009). Validity and reliability of the Screening Test for Alzheimer’s Disease with Proverbs (STADP) for the elderly. *Arq Neuropsiquiatr*, 67(3-B), 836-842.
- Saussure, F. (1916). *Cours de linguistique générale*. In Édition critique préparée par De Mauro. T. (1978), Payot.
- Schneider, G., North, B., Flügel, C., & 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>).
- Sé, E. (2011). *Interpretação de provérbios por sujeitos com doença de Alzheimer em fase inicial*. (Tese doutoramento não publicada). Universidade Estadual de Campinas, Campinas.
- Sevilla Muñoz, J. (1993). Las paremias españolas: clasificación, definición y correspondencia francesa. *Paremia*, 2, 15-20.
- Silva, J., Branco, A., Castro, S., & Reis, R. (2010a). Out-of-the-Box Robust Parsing of Portuguese. In *Proceedings of the 9th International Conference on the Computational Processing of Portuguese (PROPOR’10)*, pp. 75-85.
- Silva, J., Branco, A., Castro, S., & Reis, R. (2010b, April 2010). Out-of-the-Box Robust Parsing of Portuguese. In T. Pardo, A. Branco, K. Aldebaro, R. Vieira, & V. Strube de Lima (Eds.), *Computational Processing of Portuguese. Proceedings*

- of the 9th International Conference on the Computational Processing of Portuguese (PROPOR'2010) (Vol. 6001). Porto Alegre RS, Brazil: Springer, pp. 75-85.
- 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, LNCS, volume 8854, Springer, pp. 239-248.
- Simões, G. (2000). *Dicionário de expressões populares portuguesas: Arcaísmos, regionalismos, calão e gíria, ditos, frases feitas, lugares-comuns, aportuguesamentos, estrangeirismos e curiosidades de linguagem*. (2ª ed.), Lisboa: Publicações Dom Quixote.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.
- Siqueira, M., Marques, D., & Gibbs, R. (2016). Metaphor-related figurative language comprehension in clinical populations: a critical review. *Scripta* 20(40), pp. 36-60.
- Small, S., Flores, D., & Noll, D. (1998). Different neural circuits subserve reading before and after therapy for acquired dyslexia. *Brain and Language*, 62, pp. 298-308.
- Steyer, K. (2015). Proverbs from a Corpus Linguistic Point of View. In Hrisztova-Gotthardt, H. and Aleksa Varga, M. (Eds.) *Introduction to Paremiology. A Comprehensive Guide to Proverb Studies*. De Gruyter, Berlin, Boston, pp. 206-228.
- Taylor, A. (1975). Selected writings on proverbs. In W. Mieder (Ed.), Helsinki, FF Communications, vol. XCI, 216.
- Taylor, W. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 30, pp. 415-433.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work (Studies in Ccorpus Linguistics 6)*. Amsterdam / Philadelphia: John Benjamins.
- Vale, O. (2001). *Expressões Cristalizadas do Português do Brasil: uma proposta de tipologia*. (Tese doutoramento não publicada) UNESP, Araraquara.
- Vas, A., Spence, J., Eschler, B., & Chapman, S. (2016). Sensitivity and specificity of abstraction using gist reasoning measure in adults with traumatic brain injury. *Journal of Applied Biobehavioral Research* 21(4), pp. 216-224.
- Vicente, A. (2013). *LexMan: um Segmentador e Analisador Morfológico com Transdutores*. (Dissertação mestrado não publicada). Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa.
- Zuluaga, A. (1975a). Estudios generativo-transformativistas de las expresiones idiomáticas. *Boletín del Instituto Caro y Cuervo, Thesaurus*, XXX (1), pp. 1-48.
- Zuluaga, A. (1975b). La fijación fraseológica. *Thesaurus*, XXX, pp. 225-248.
- Zuluaga, A. (1980). *Introducción al estudio de las expresiones fijas*. Frankfurt: Peter Lang.

Anexos

.1 Lista de *stopwords*

.	em	delas	no	esta
,	entre	deste	na	estes
:	excepto	desta	nos	estas
;	mediante	destes	nas	esse
?	para	destas	nesse	essa
!	perante	desse	nessa	esses
...	por	dessa	nesses	essas
"	salvo	desses	nessas	isso
"	sem	dessas	neste	aquele
«	sob	daquele	nesta	aquela
»	sobre	daquela	nestes	aqueles
(trás	daqueles	nestas	aquelas
)	o	daquelas	nisto	aquilo
{	a	daqui	nisso	mesmo
}	os	dali	naquilo	mesma
[as	dacolá	noutro	mesmos
]	um	daquém	noutra	mesmas
-	uma	dalém	noutros	próprio
-	uns	donde	noutras	própria
/	umas	dentre	num	próprios
a	ao	doutro	numa	próprias
o	à	doutra	nuns	outro
&	aos	doutros	numas	outra
*	às	doutras	naquele	outro
+	àquele	doutrem	naquela	outras
=	àquela	disso	naqueles	outrem
ante	àqueles	disto	naquelas	algum
após	àquelas	daquilo	pelo	alguma
até	àquilo	dum	pela	alguns
com	do	duma	pelos	algumas
conforme	da	duns	pelas	alguém
contra	dos	dumas	ele	ninguém
consoante	das	estoutro	ela	meu
de	dele	estoutra	eles	minha
desde	dela	estoutros	elas	meus
durante	deles	estoutras	este	minhas

nosso	quanta	to	eh	sim
nossa	quantos	ta	e	já
nossos	quantas	tos	que	também
nossas	qual	tas	se	depois
vosso	quais	lhe	como	antes
vossa	quem	lho	quanto	agora
vossos	porquê	lha	mas	então
vossas	lo	lhos	porém	apenas
seu	la	lhas	ou	ainda
sua	los	mim	quando	jamais
seus	las	ti	logo	tanto
suas	vo	si	enquanto	tão
teu	vos	<ser>	ou	talvez
tua	me	<estar>	porque	senão
teus	mo	<dever>	contudo	excepto
tuas	ma	adeus	pois	eis
qualquer	mos	ai	portanto	através
quaisquer	mas	oh	ora	facilmente
quanto	te	ah	não	raramente

.2 Critérios para agrupar variantes de provérbios sob uma mesma *Unidade Paremiológica*

Para agrupar variantes de provérbios sob uma mesma UP, deverão ser seguidos os seguintes critérios (em conjunto e sucessivamente):

1º Critério: Apesar da variação, o significado global do provérbio deve ser idêntico;

2º Critério: Deve envolver os mesmos elementos lexicais-chave – os que se repetem em todos os provérbios;

Critério adicional: Ainda assim, considera-se variante se um provérbio apresentar variação lexical dos elementos-chave, mas se as expressões tiverem um significado global idêntico e puderem ser usadas num mesmo contexto discursivo.

Exemplos

(1) Filho de peixe sabe nadar

(2) Quem sai aos seus não degenera

Embora os provérbios (1) e (2) possam eventualmente ser considerados variantes semânticas (significado global idêntico), considera-se que se trata de dois provérbios **distintos**, uma vez que **não** partilham nenhuns elementos lexicais-chave.

(3a) Roma e Pavia não se fizeram num dia

(3b) Não se fez Roma e Pavia num dia

(3c) Roma não se fez num dia

(3d) Não se fez Roma em um dia

(3e) França e Roma não se fez num dia

Neste caso, trata-se de variantes do mesmo provérbio, apenas diferentes na ordem dos elementos-chave por que são formadas. No último caso, uma das palavras-chave varia mas há ainda suficientes palavras-chave em comum, podendo invocar-se o critério adicional.

.3 Avaliação do OCR (*Optical Character Recognition*)

Para a constituição da base de dados, procedeu-se à digitalização do conteúdo integral de quatro coletâneas de provérbios (Costa 1999, Machado 2011, Moreira 1996, Parente 2005). Ao digitalizar o conteúdo das coletâneas (impressas) em texto editável (i.e. um texto legível por máquinas – *machine readable*), ocorrem, geralmente, alguns erros⁴. Esses erros, se não forem corrigidos manualmente, poderão interferir nos passos de análise subsequentes. Impunha-se, pois, determinar a taxa de erros gerados pelo OCR e assim aferir a qualidade do texto digitalizado. Para tal, seguimos o seguinte procedimento:

1. Seleccionámos uma amostra a ser testada;
2. Determinámos o número e tipo de erros encontrados;
3. Calculámos a taxa de erro de cada uma das coletâneas;
4. Calculámos a média da taxa de erro das 4 coletâneas.

As coletâneas usadas para construir a base de dados de provérbios deste trabalho têm cada uma um número variável de páginas. Numa primeira fase, seleccionámos apenas as páginas que continham provérbios, descartando assim toda a informação relativa à ficha técnica, ao prefácio, à bibliografia, entre outras. De seguida, seleccionámos cerca de 5% das páginas de cada coletânea, das listagens de provérbios já digitalizadas. Depois, copiámo-las para um editor de texto (MS-WordTM), a fim de verificar o número de erros, bem como a sua tipologia. Na tabela seguinte, apresentamos as informações relativas a cada uma das coletâneas em análise.

Tabela 2: *Avaliação OCR*

Coletânea	nº páginas (5%)	nº palavras	nº erros	taxa de erro
Parente (2005)	37	20.283	3	0,0001
Machado (2011)	31	13.935	34	0,0024
Costa (1999)	25	14.502	7	0,0005
Moreira (1996)	15	6.027	10	0,0017
total	108	54.747	54	0,0047

O número de erros encontrados por coletânea é variável, como seria de esperar, pois concorrem para a qualidade do OCR o tipo de papel do texto impresso, a fonte utilizada, o espaçamento entre linhas, a paginação, etc.

Na digitalização de Machado (2011) encontrámos mais erros do que nas restantes coletâneas (34). Em todas as coletâneas, aliás, a maior parte dos erros está relacionado com problemas de translineação. Note-se que 20 dos erros de Machado (2011) são problemas de translineação com hífen (*más- -línguas* por *más-línguas*) ou a inserção de espaços (*af ectam* por *afectam*), havendo alguns erros de acentuação (*arruína* por *arruína* e *agüenta* por *aguenta*). Em Costa (1999) 6 dos 7 erros encontrados são erros de translineação.

⁴No processo de digitalização usou-se o scanner Fujitsu ScanSnap S1500M e o software OCR Scan to Word.

O número de erros encontrado por coletânea não está diretamente relacionado com o número de palavras de cada uma. Em Parente (2005) foram encontrados 3 erros num total de 20.283 palavras (taxa de erro: 0,00015). Em Moreira (1996) foram encontrados 10 erros nas 6.027 palavras da amostra (taxa de erro: 0,0016).

Embora a correção manual possa ter deixado passar algum erro, é, pois, esperável que a qualidade do material textual recolhido e digitalizado seja adequada às tarefas a que nos propomos nesta tese, concretamente, a identificação de provérbios em textos, uma vez que num total de 54.747 palavras foi encontrado, em média, 1 erro por cada 1.000 palavras.

.4 Grafos

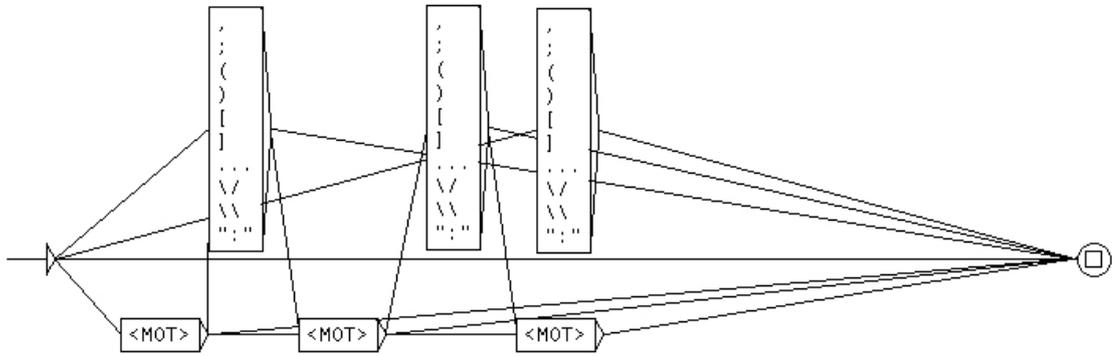


Figura 2: Grafo de inserção 1_3

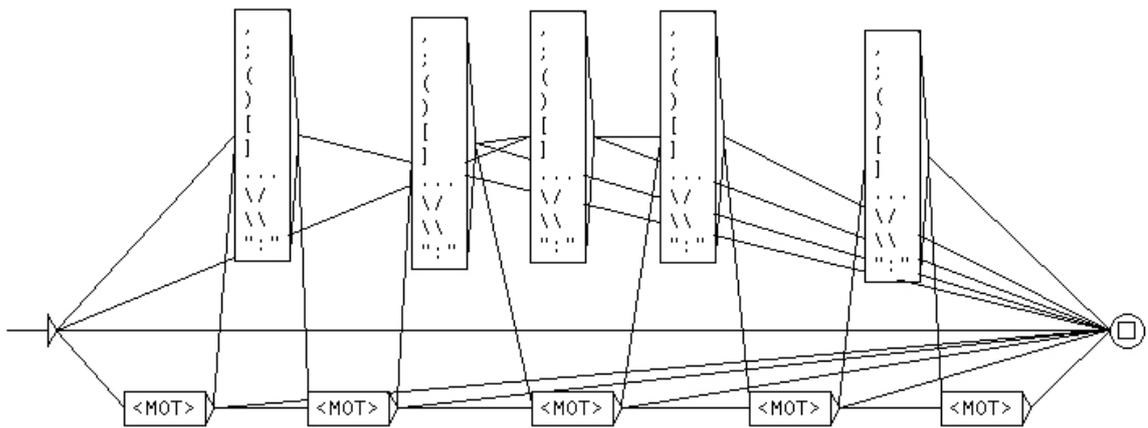


Figura 3: Grafo de inserção 1_5

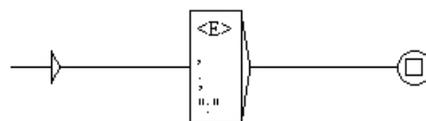


Figura 4: Grafo de pontuação e <E>

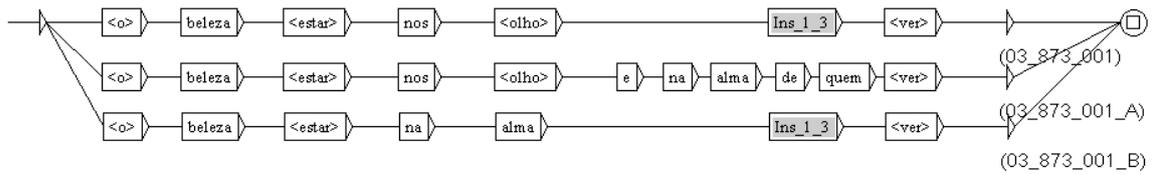


Figura 5: Grafo: *A beleza está nos olhos de quem vê*



Figura 6: Grafo: *A carne é fraca*

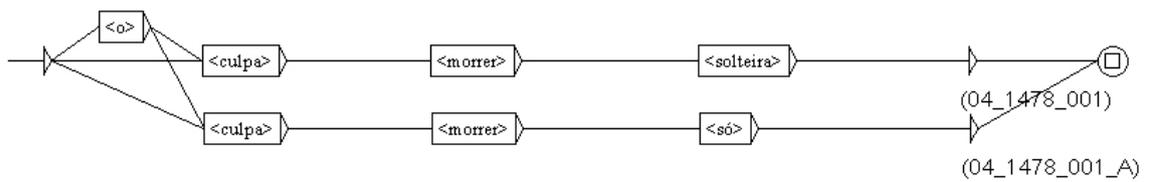


Figura 7: Grafo: *A culpa morreu solteira*

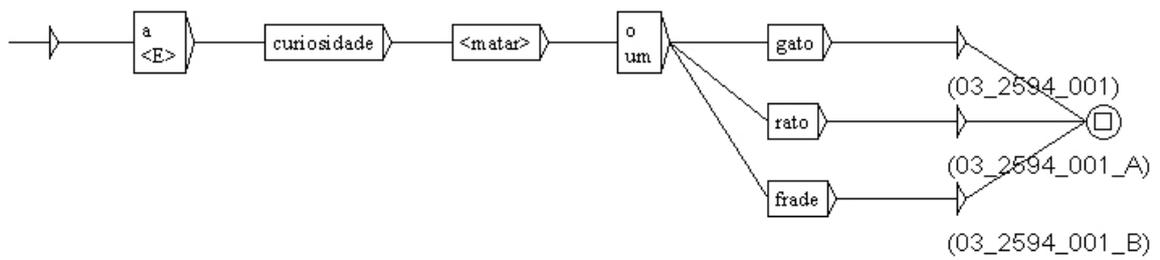


Figura 8: Grafo: *A curiosidade matou o gato*

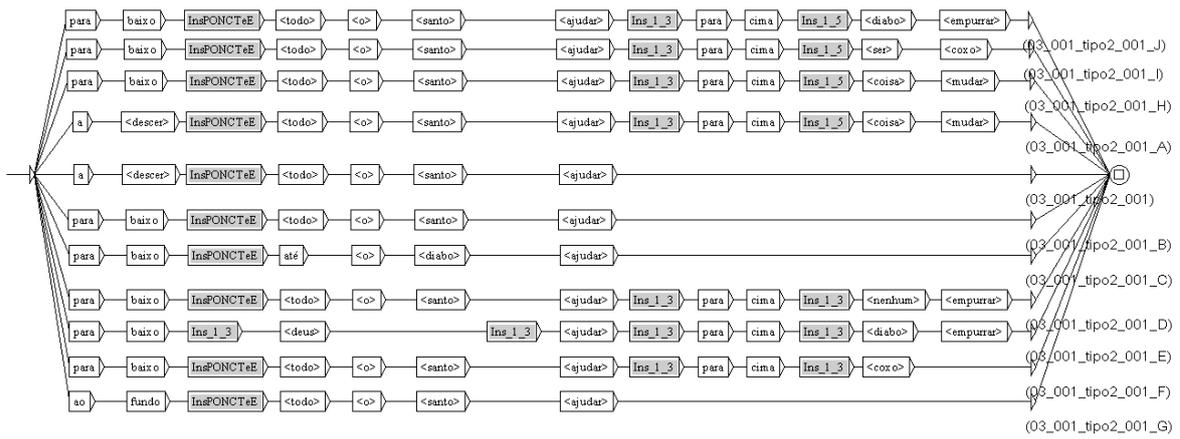


Figura 9: Grafo: Para baixo todos os santos ajudam

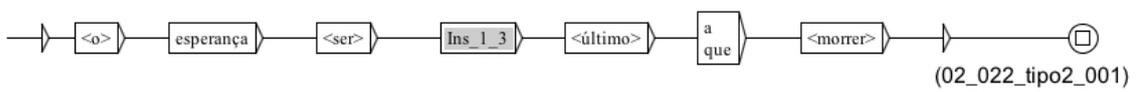


Figura 10: Grafo: A esperança é sempre a última a morrer

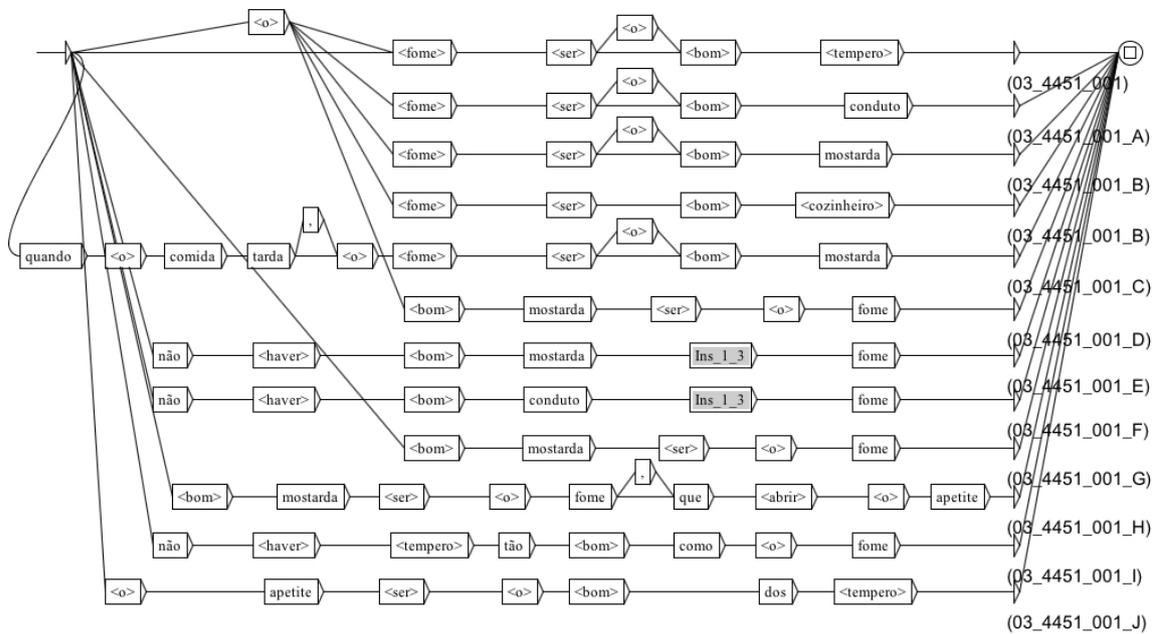


Figura 11: Grafo: A fome é o melhor tempero



Figura 12: Grafo: A intenção é que conta

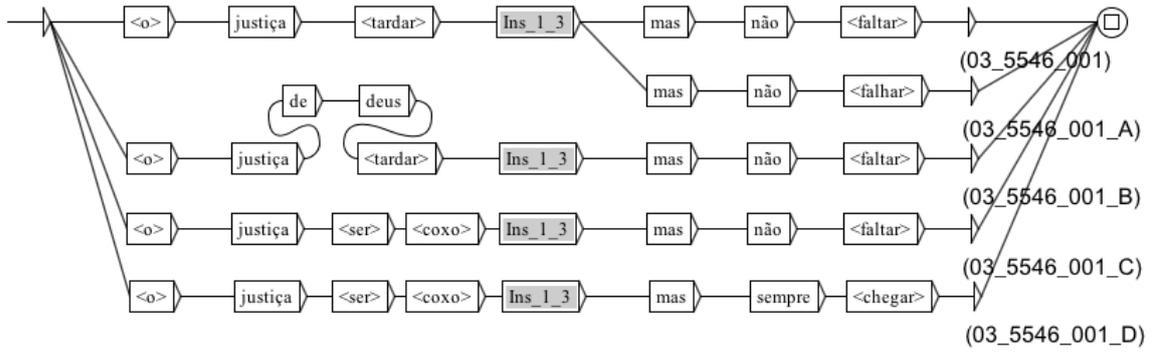


Figura 13: Grafo: A justiça tarda mas não falta

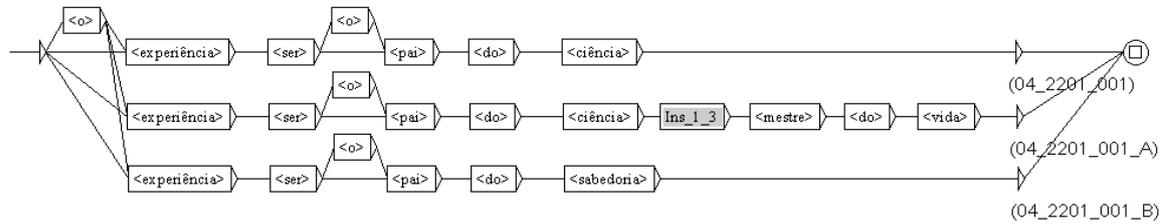


Figura 14: Grafo: A experiência é a mãe da ciência

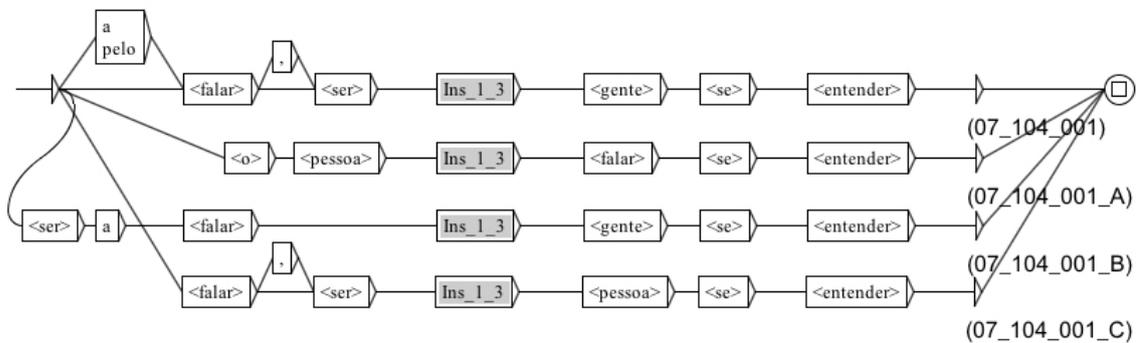


Figura 15: Grafo: A falar é que a gente se entende

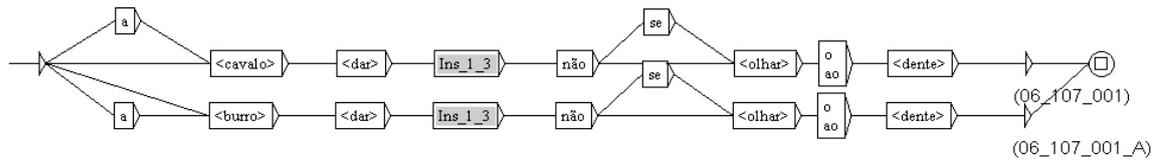


Figura 16: Grafo: A cavalo dado não se olha ao dente

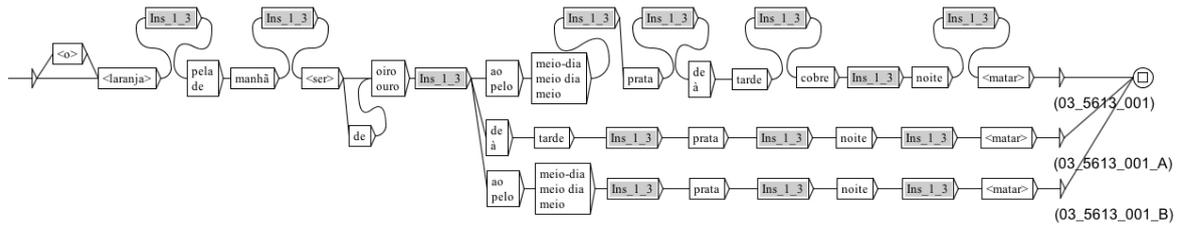


Figura 17: Grafo: A laranja pela manhã é ouro, ao meio-dia prata, à tarde cobre e à noite mata

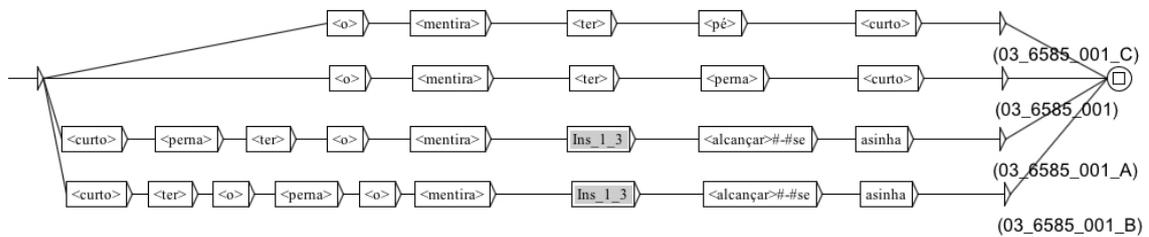


Figura 18: Grafo: A mentira tem pernas curtas

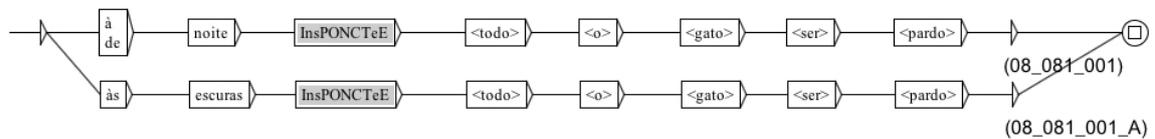


Figura 19: Grafo: À noite todos os gatos são pardos

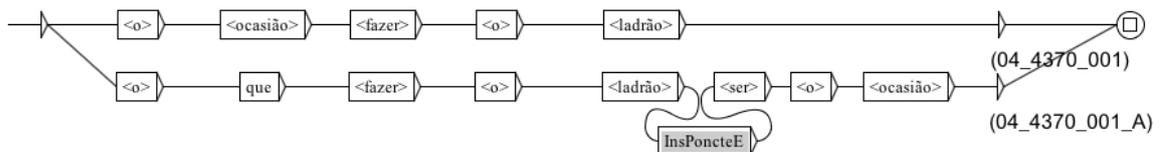


Figura 20: Grafo: A ocasião faz o ladrão

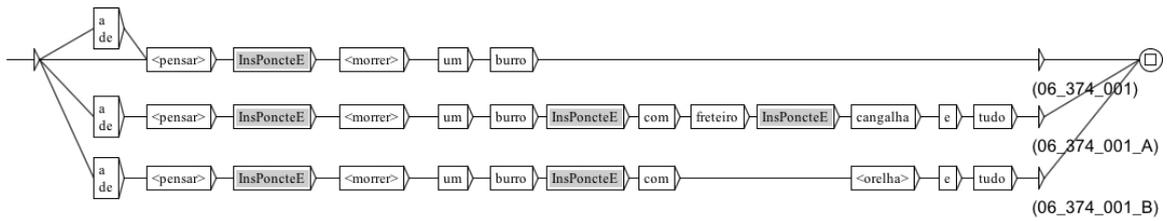


Figura 21: Grafo: A pensar morreu um burro

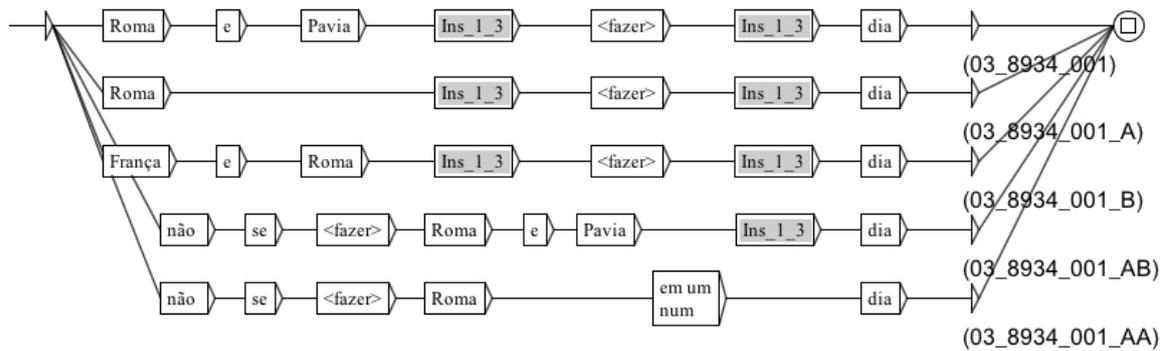


Figura 22: Grafo: Roma e Pavia não se fizeram num dia

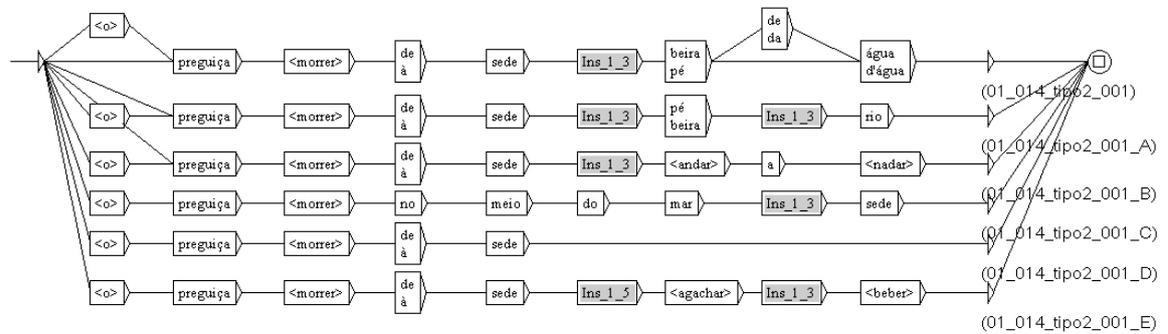


Figura 23: Grafo: A preguiça morreu à sede à beira do rio

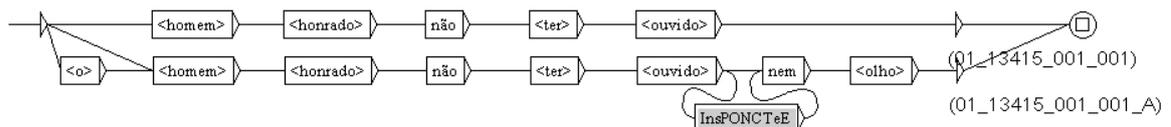


Figura 24: Grafo: A mulher honrada não tem ouvidos

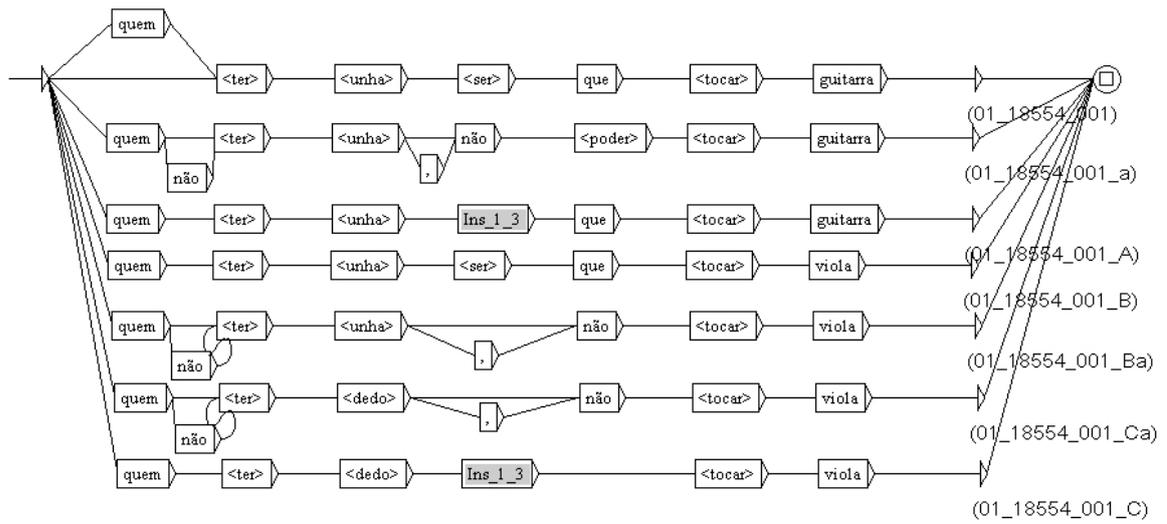


Figura 25: Grafo: Quem tem unhas é que toca guitarra

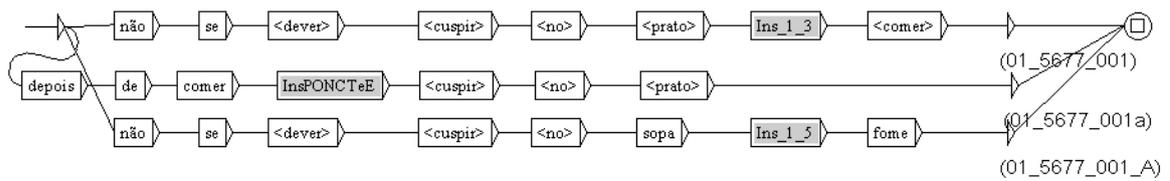


Figura 26: Grafo: Não se deve cuspir no prato onde se comeu

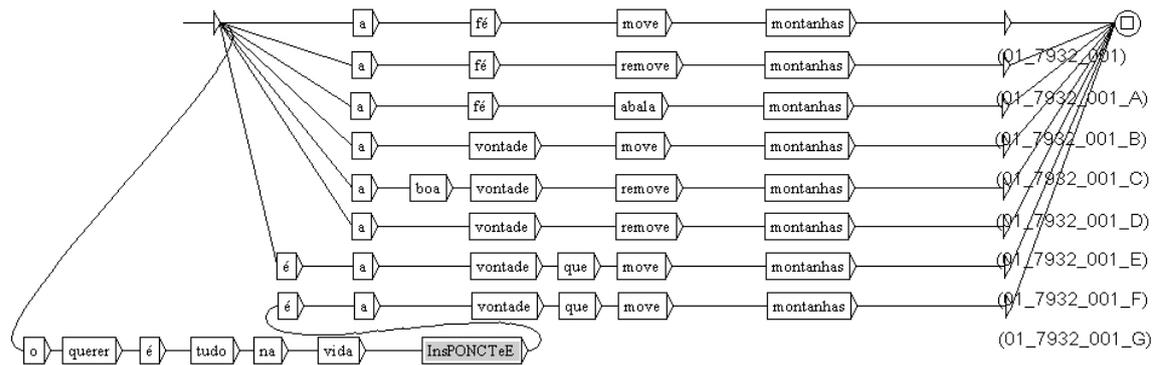


Figura 27: Grafo: A fé move montanhas

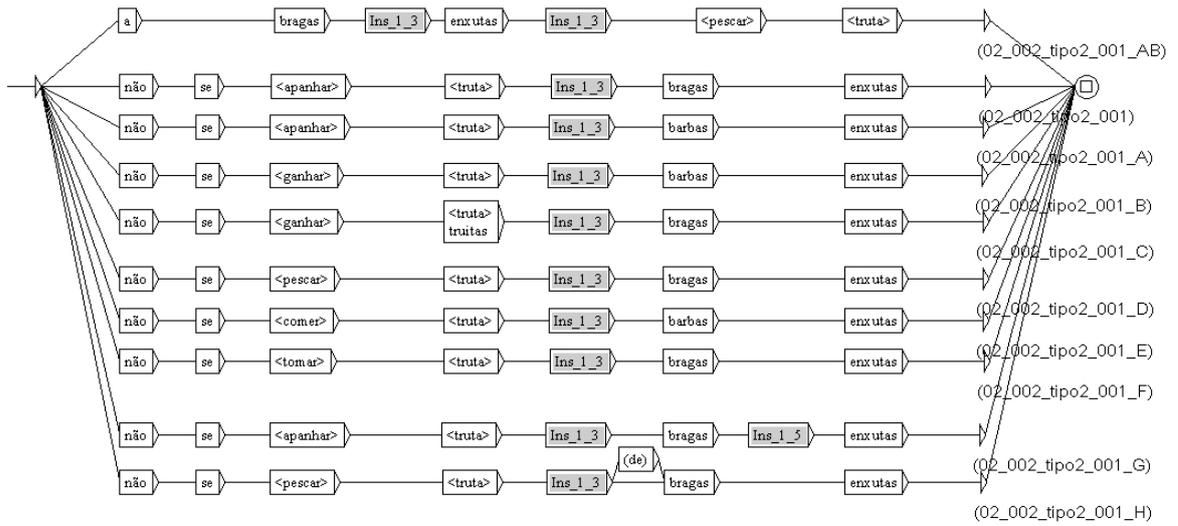


Figura 28: Grafo: Não se apanham trutas a bragas enxutas

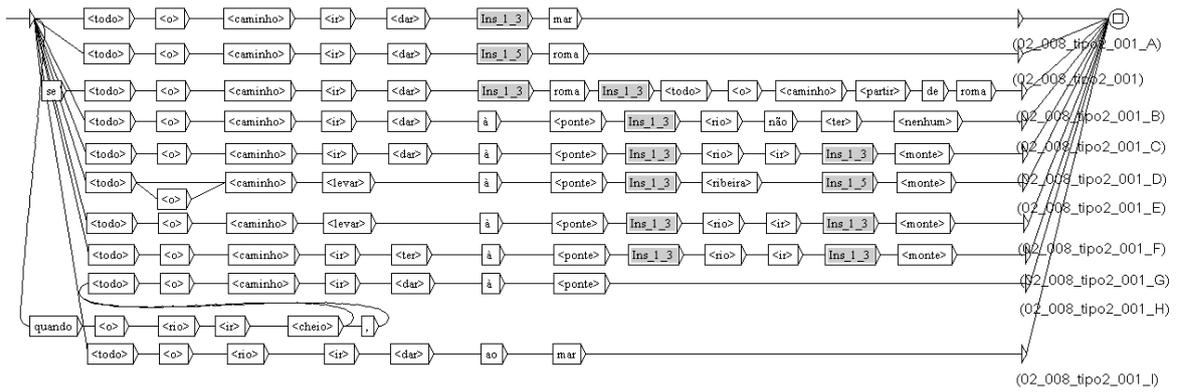


Figura 29: Grafo: Todos os caminhos vão dar a Roma

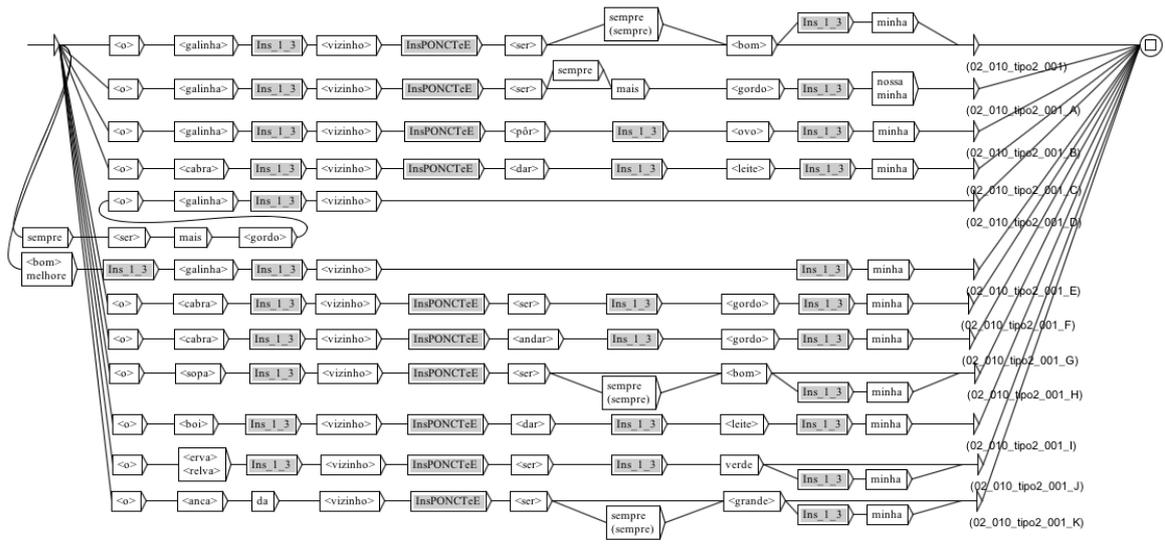


Figura 30: Grafo: *A galinha da vizinha é sempre melhor que a minha*

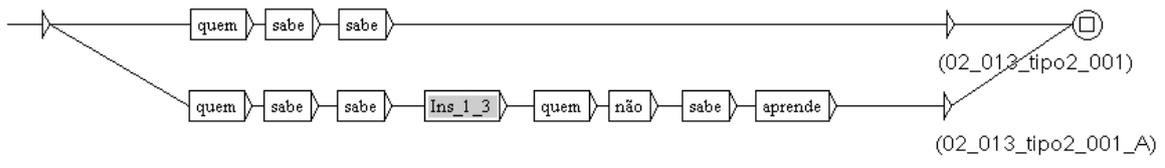


Figura 31: Grafo: *Quem sabe, sabe*

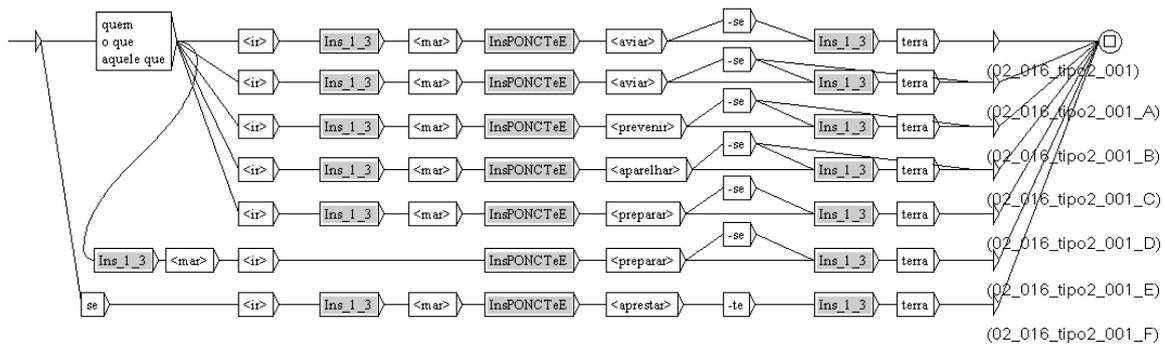


Figura 32: Grafo: *Quem vai para o mar avia-se em terra*

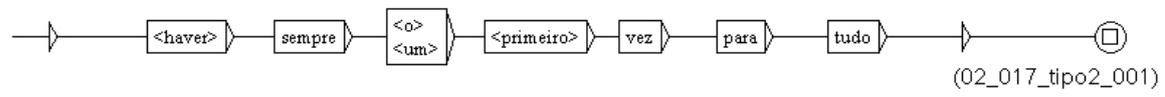


Figura 33: Grafo: *Há sempre uma primeira vez para tudo*

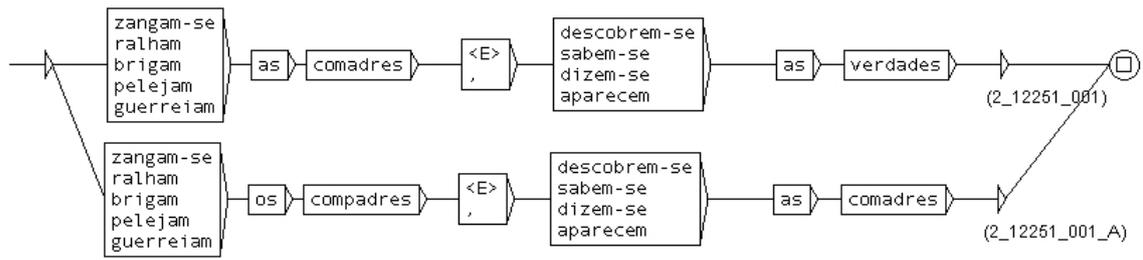


Figura 34: Grafo: *Zangam-se as comadres, descobrem-se as verdades*

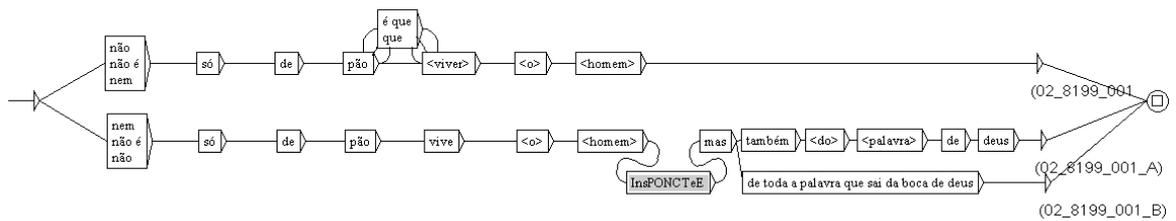


Figura 35: Grafo: *Nem só de pão vive o homem*

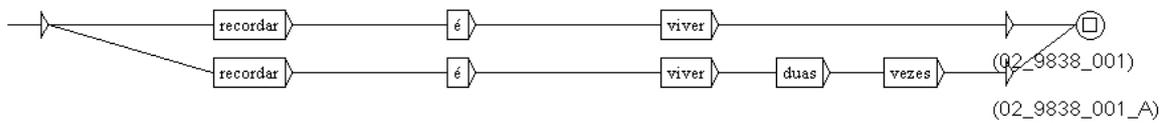


Figura 36: Grafo: *Recordar é viver*

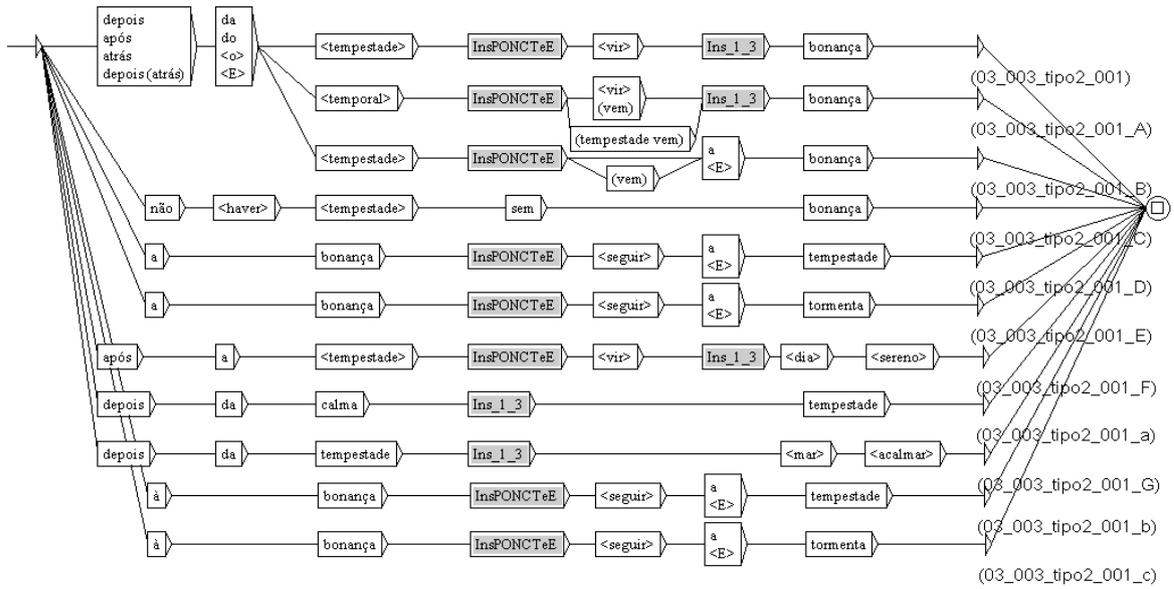


Figura 37: Grafo: Depois da tempestade vem a bonança

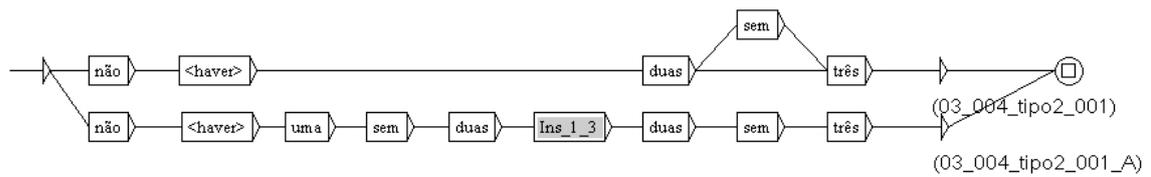


Figura 38: Grafo: Não há duas sem três

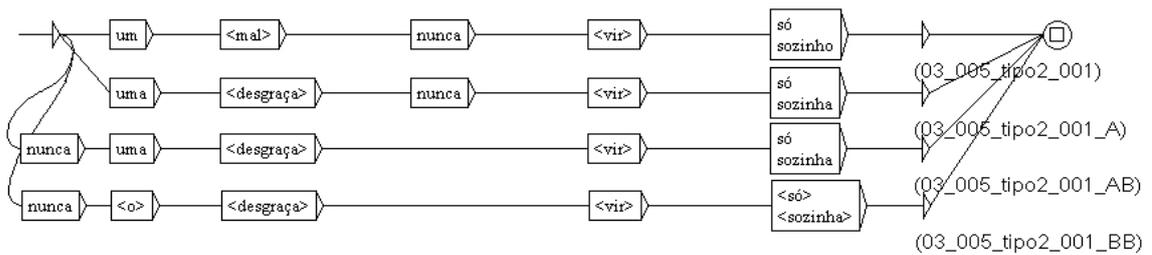


Figura 39: Grafo: Um mal nunca vem só

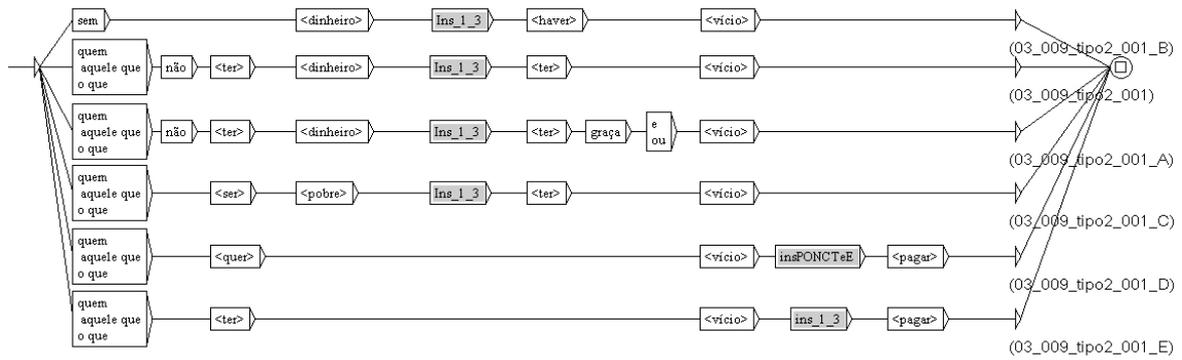


Figura 40: Grafo: Quem não tem dinheiro não tem vícios

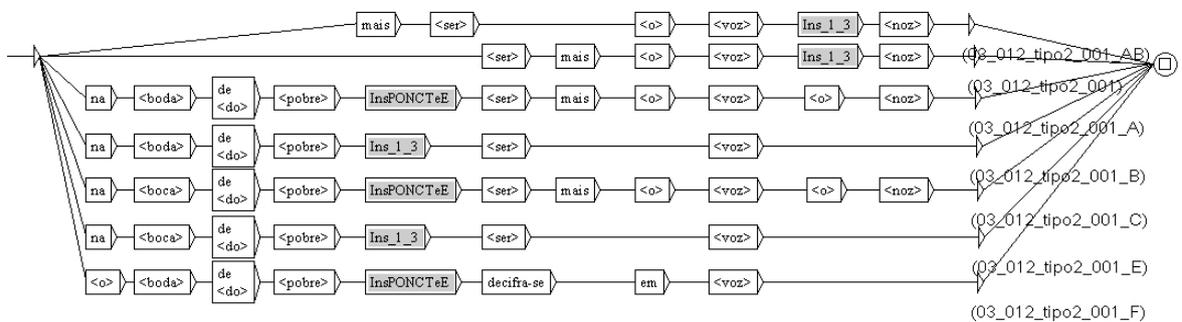


Figura 41: Grafo: São mais as vozes do que as nozes

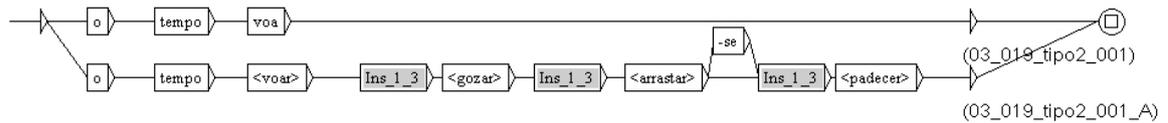


Figura 42: Grafo: O tempo voa

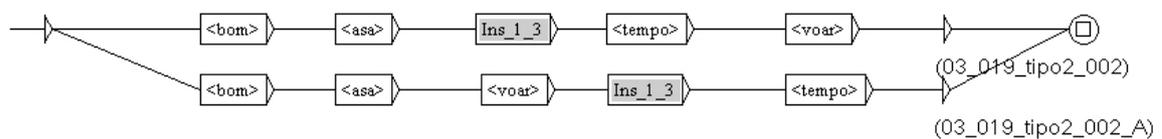


Figura 43: Grafo: Boa asa com todo o tempo voa

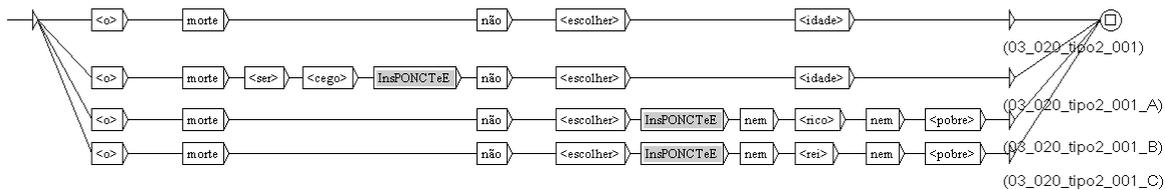


Figura 44: Grafo: A morte não escolhe idades

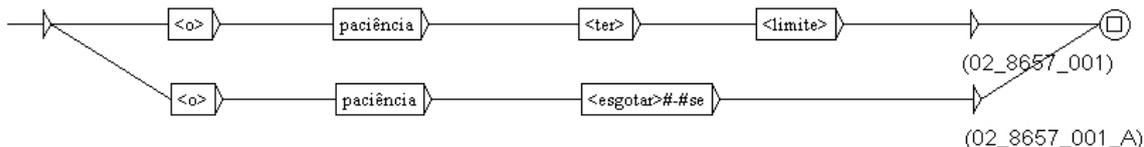


Figura 45: Grafo: A paciência tem limites

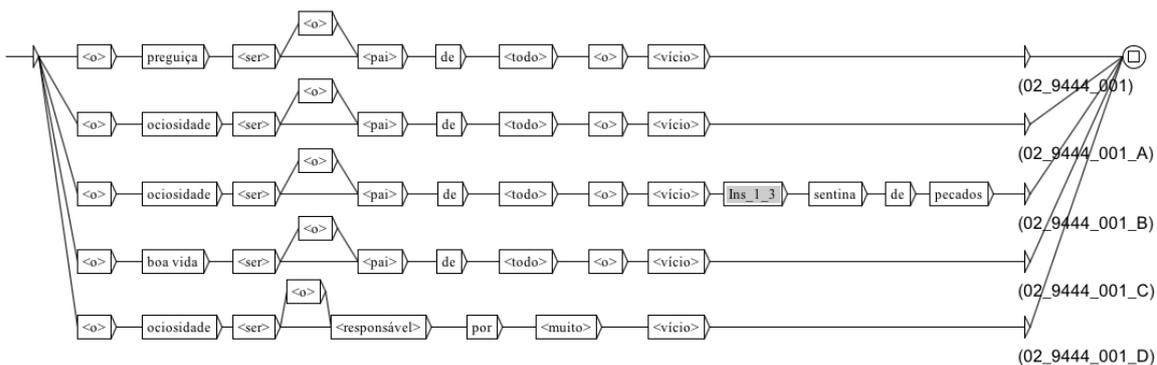


Figura 46: Grafo: A preguiça é a mãe de todos os vícios

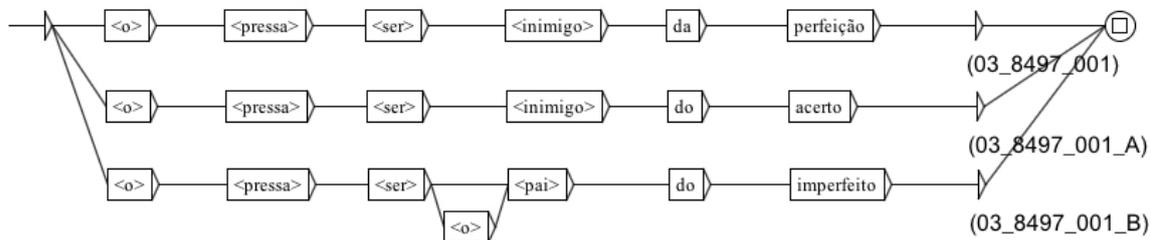


Figura 47: Grafo: A pressa é inimiga da perfeição

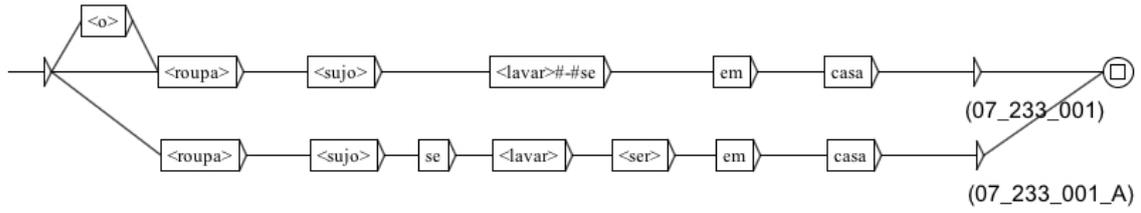


Figura 48: Grafo: A roupa suja lava-se em casa

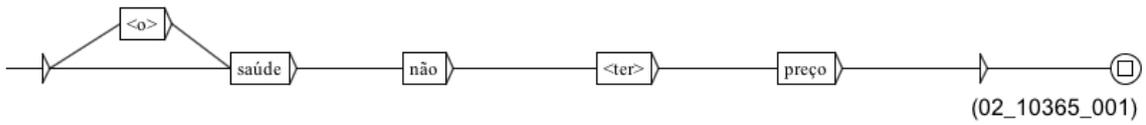


Figura 49: Grafo: A saúde não tem preço

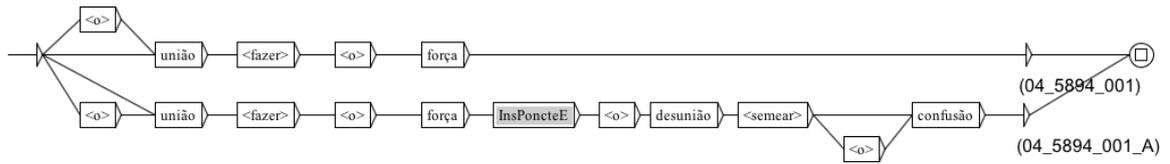


Figura 50: Grafo: A união faz a força

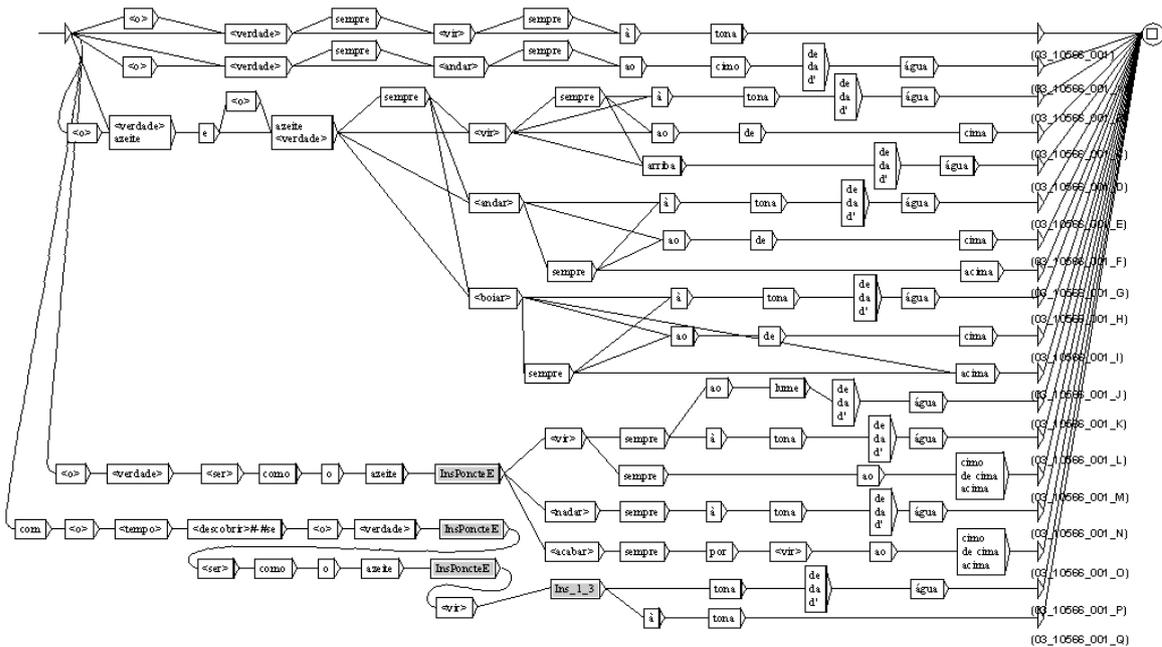


Figura 51: Grafo: A verdade sempre vem à tona

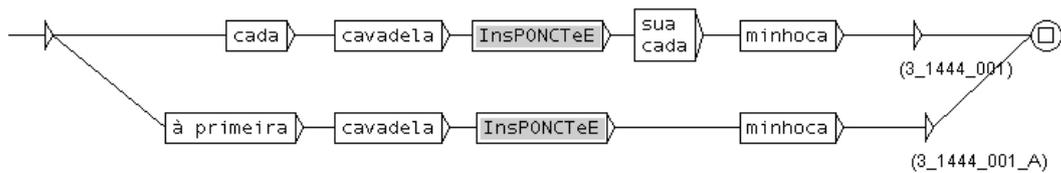


Figura 52: Grafo: À primeira cavadela minhoca

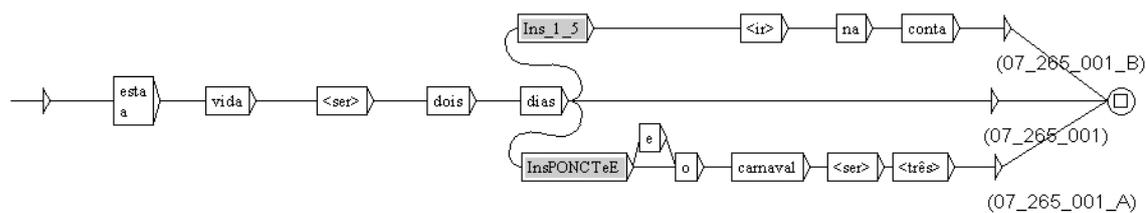


Figura 53: Grafo: A vida são dois dias

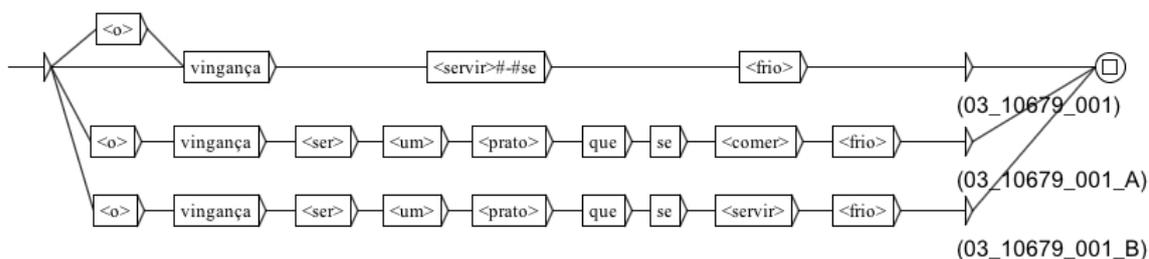


Figura 54: Grafo: A vingança serve-se fria

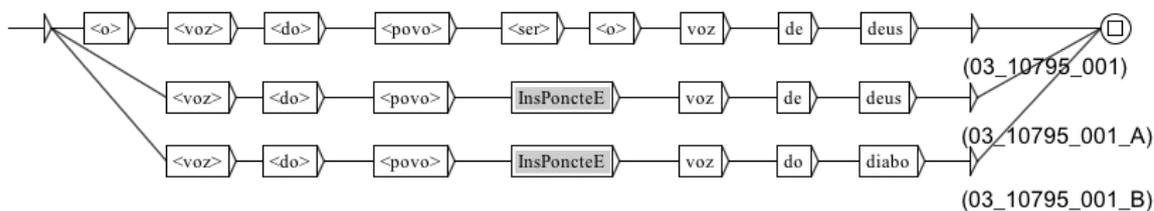


Figura 55: Grafo: A voz do povo é a voz de Deus

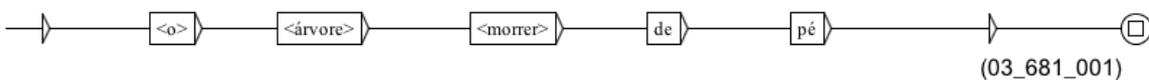


Figura 56: Grafo: As árvores morrem de pé

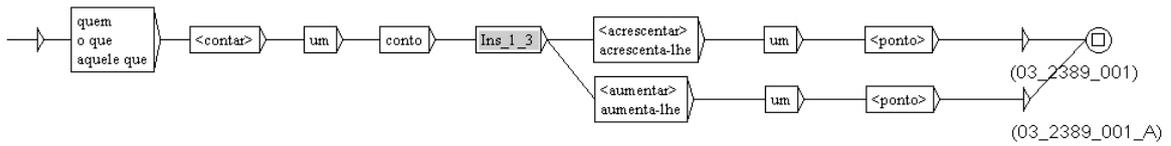


Figura 57: Grafo: Quem conta um conto acrescenta um ponto

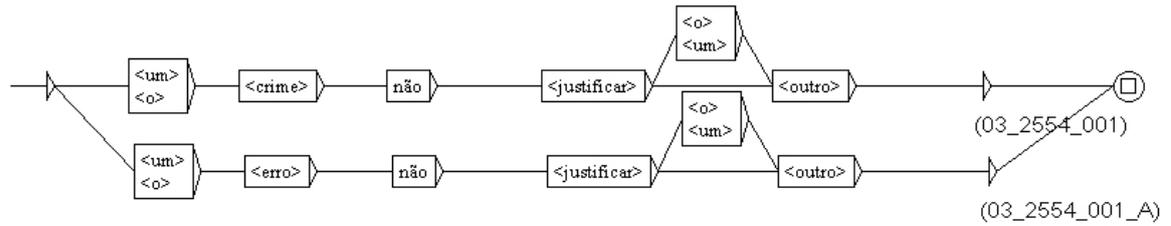


Figura 58: Grafo: Um crime não justifica outro

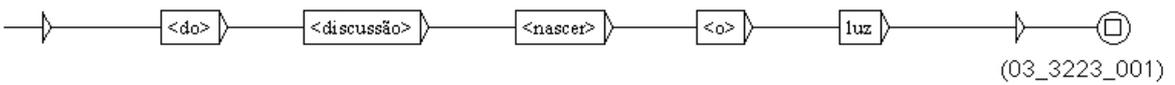


Figura 59: Grafo: Da discussão nasce a luz

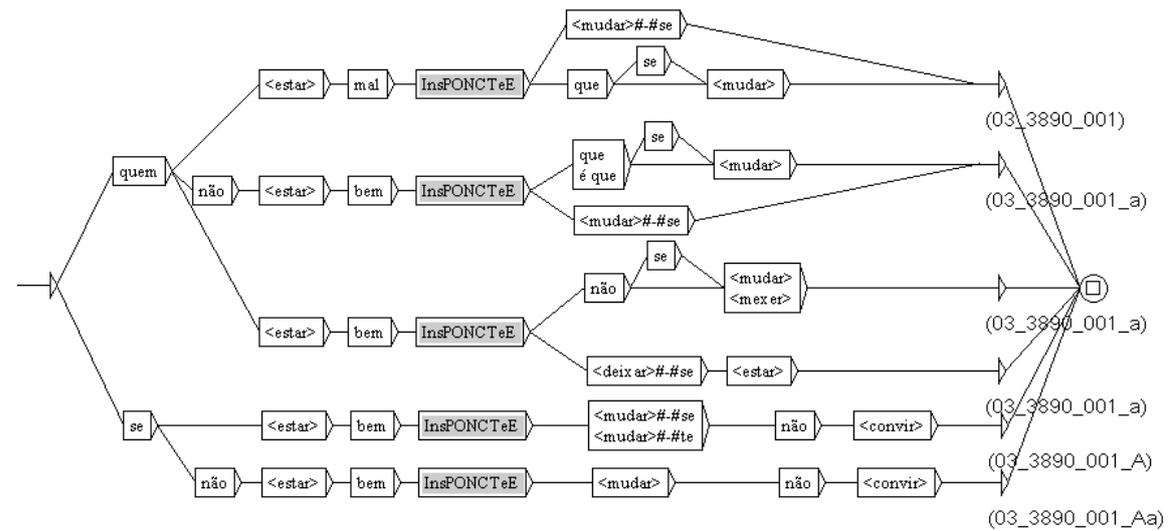


Figura 60: Grafo: Quem está mal que se muda

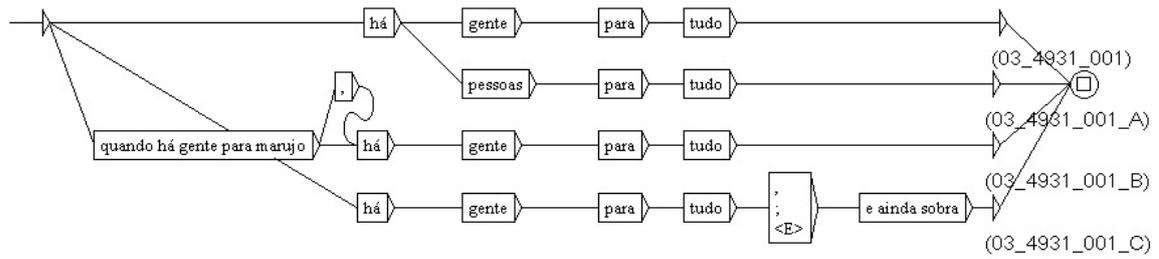


Figura 61: Grafo: Há gente para tudo

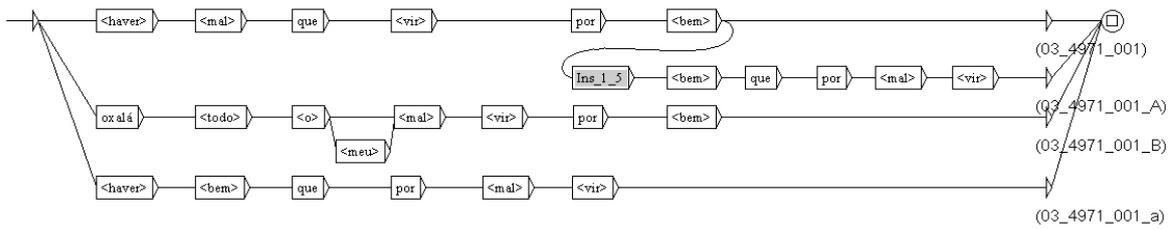


Figura 62: Grafo: Há males que vêm por bem

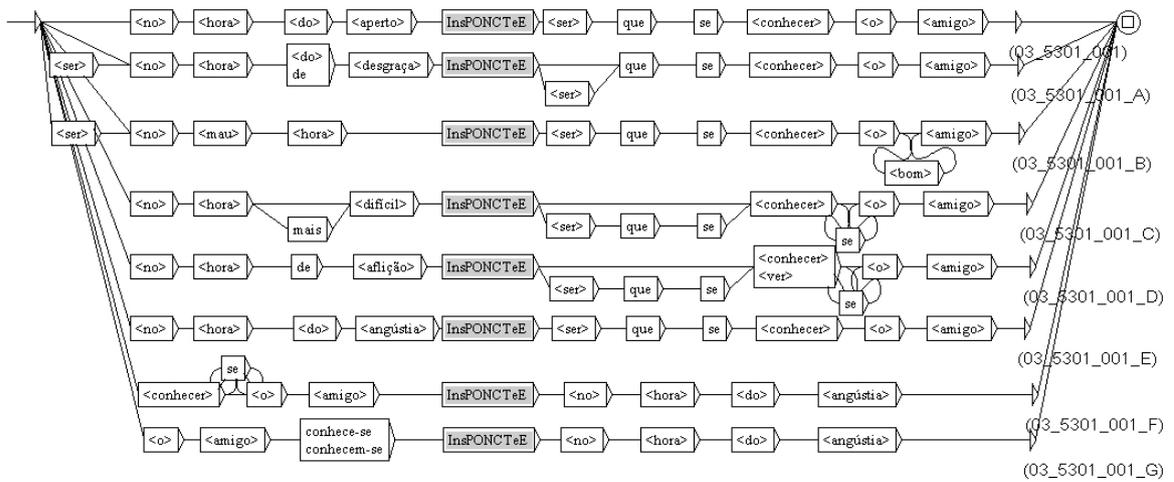


Figura 63: Grafo: Na hora do aperto é que se conhecem os amigos

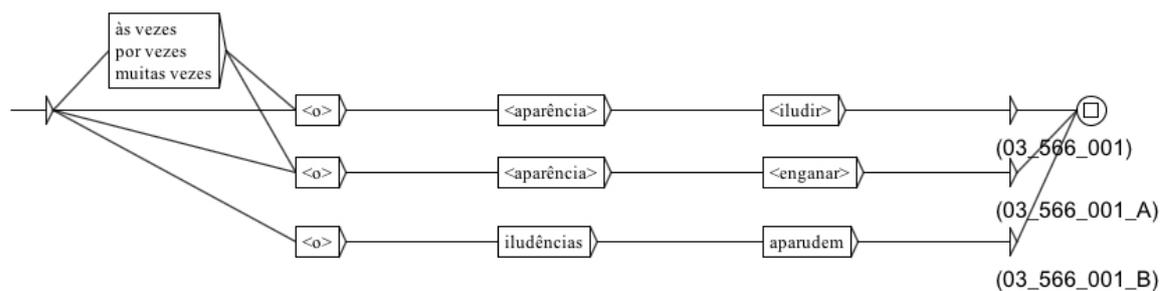


Figura 64: Grafo: As aparências iludem

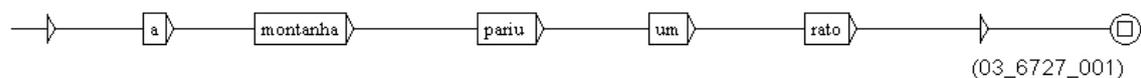


Figura 65: Grafo: A montanha pariu um rato



Figura 66: Grafo: Um santo não pode estar em dois altares

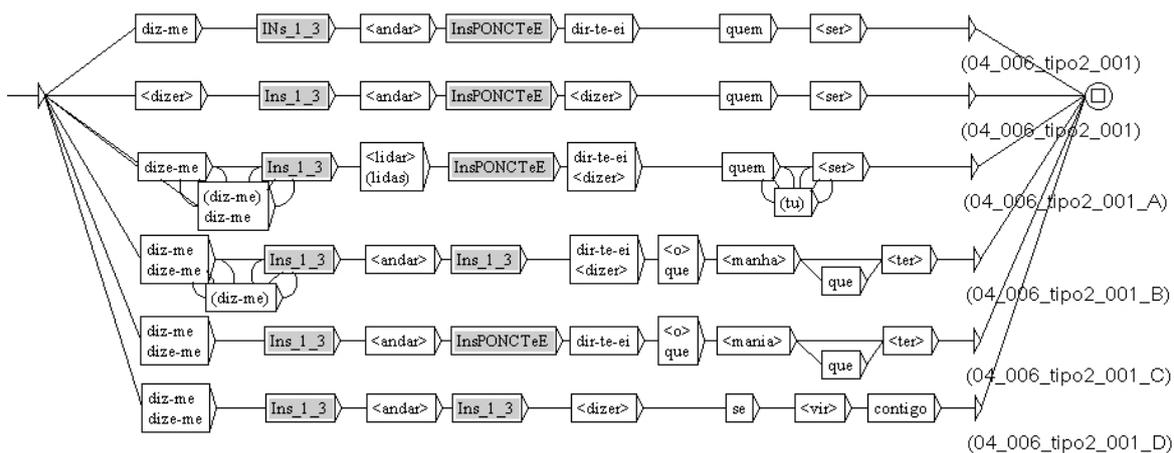


Figura 67: Grafo: Diz-me com quem andas dir-te-ei quem és

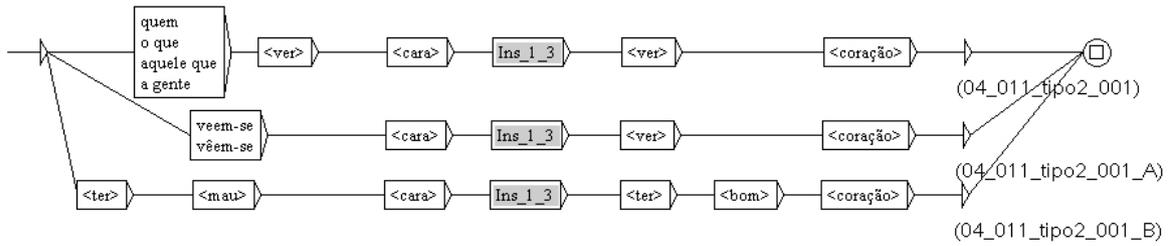


Figura 68: Grafo: Quem vê caras não vê corações

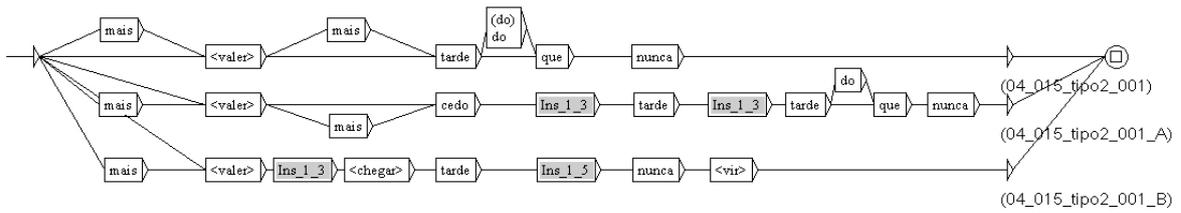


Figura 69: Grafo: Mais vale tarde do que nunca

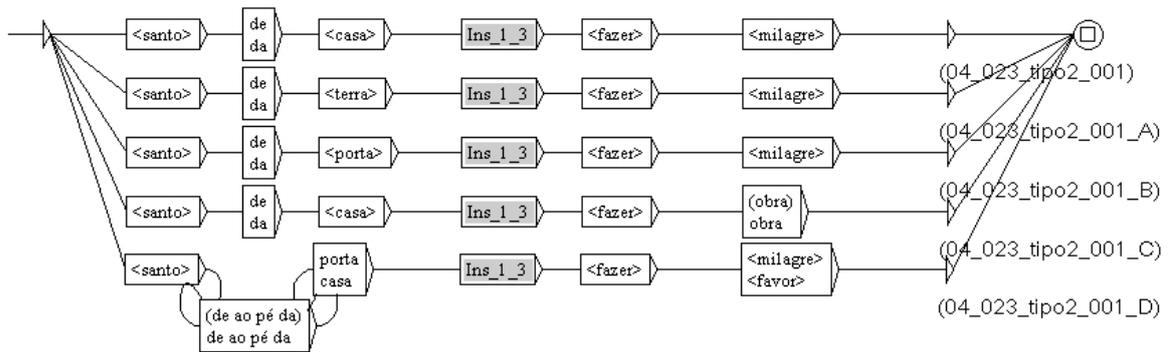


Figura 70: Grafo: Santos da casa não fazem milagres

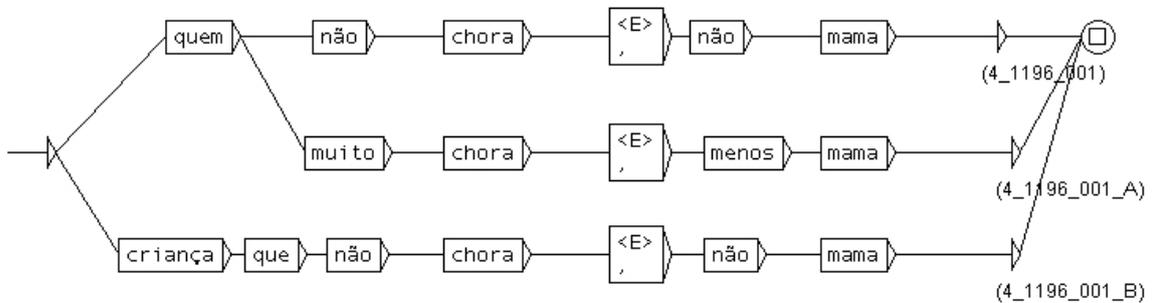


Figura 71: Grafo: Quem não chora não mama

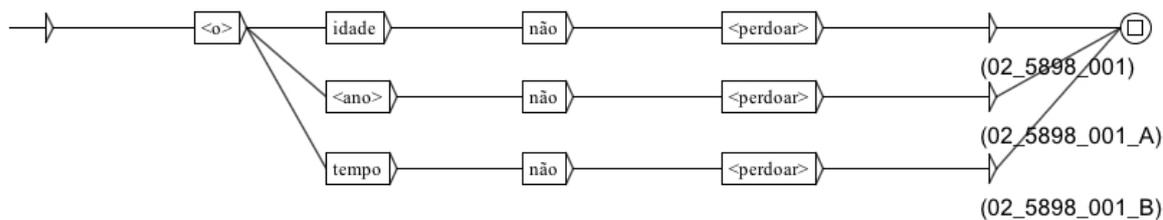


Figura 72: Grafo: A idade não perdoa

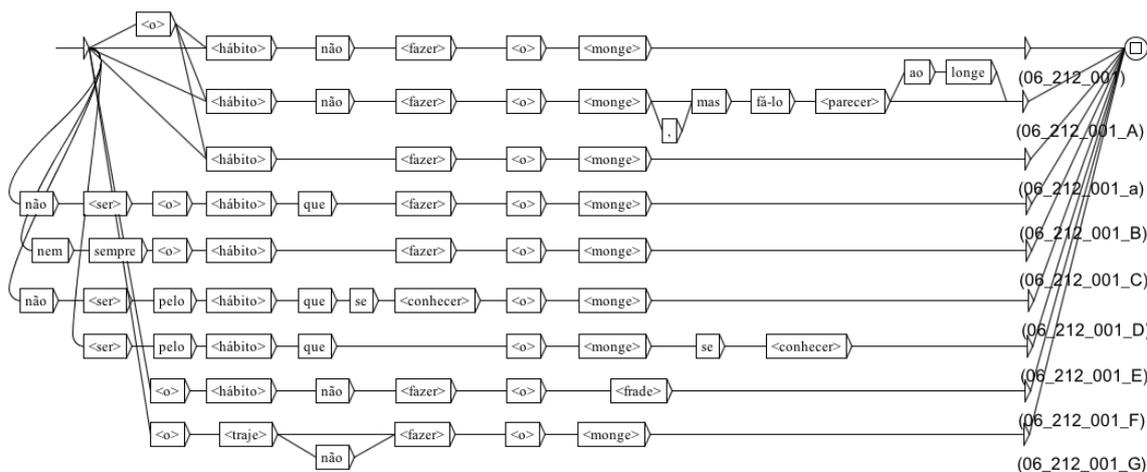


Figura 73: Grafo: O hábito não faz o monge

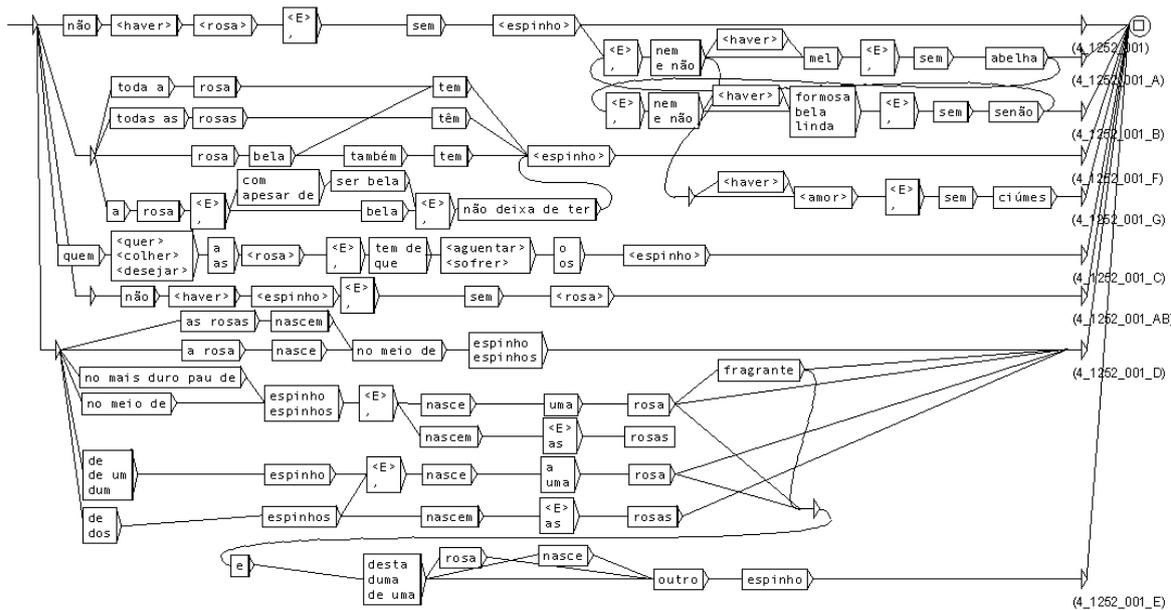


Figura 74: Grafo: Não há rosas sem espinhos

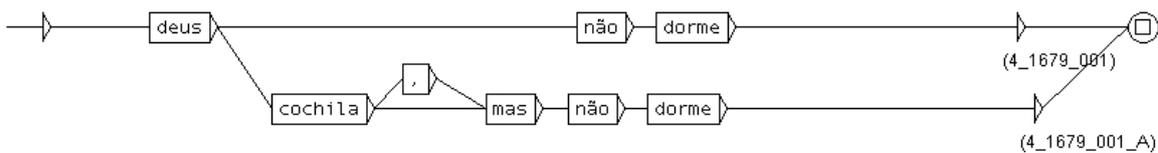


Figura 75: *Grafo: Deus não dorme*

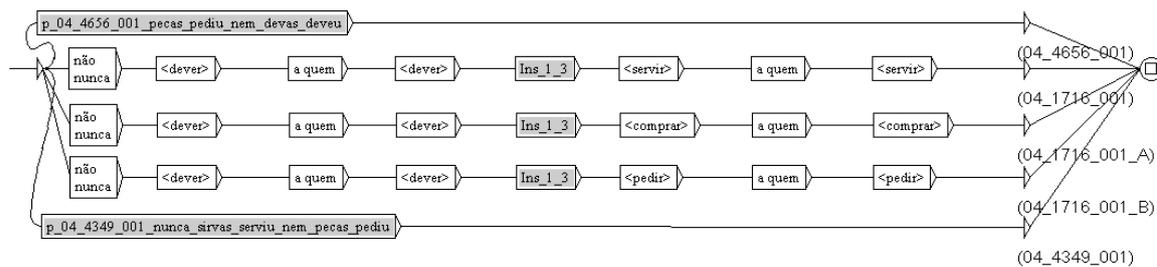


Figura 76: *Grafo: Não devas a quem deveu nem sirvas a quem serviu*

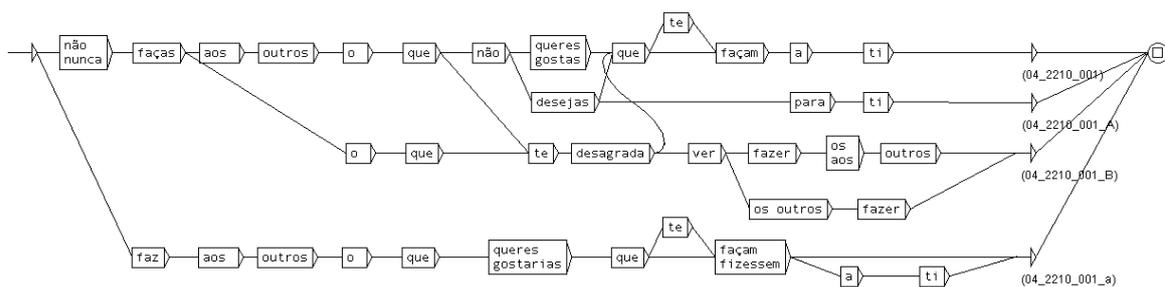


Figura 77: *Grafo: Não faças aos outros o que não gostas que te façam a ti*

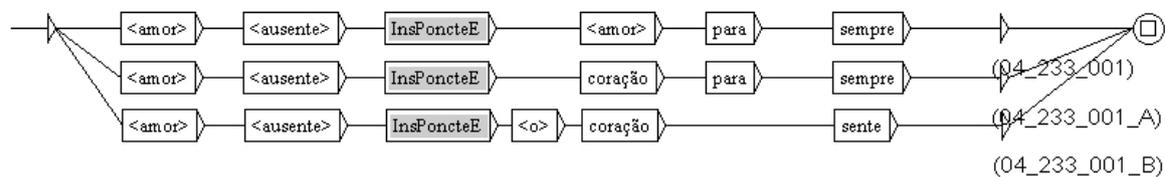


Figura 78: *Grafo: Amor ausente amor para sempre*

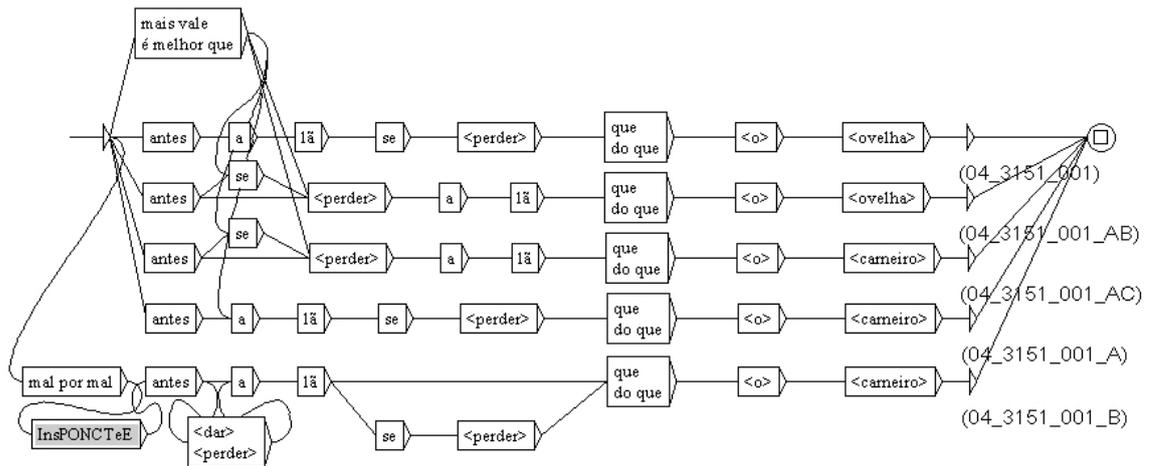


Figura 79: Grafo: Antes se perca a lã que a ovelha

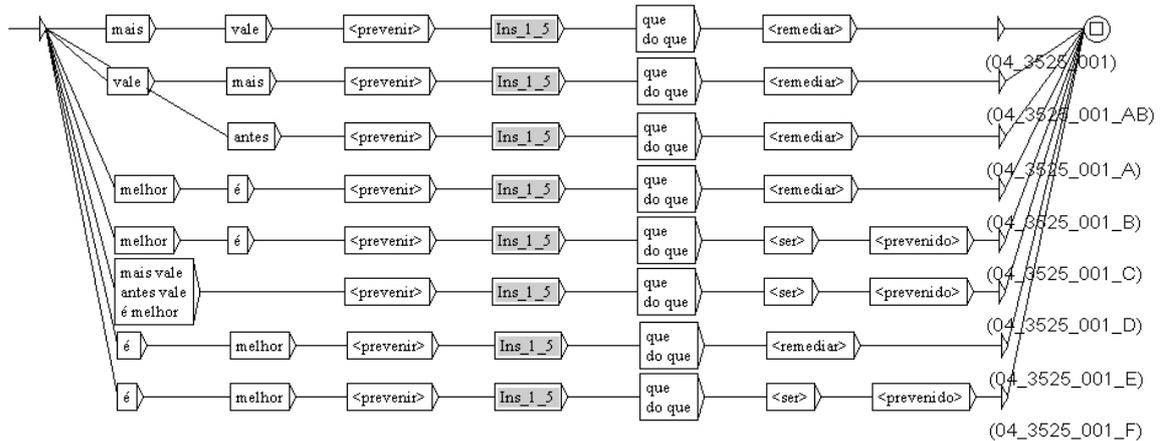


Figura 80: Grafo: Mais vale prevenir que remediar

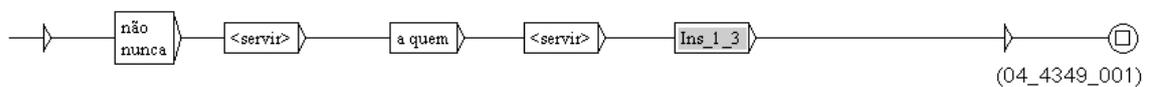


Figura 81: Grafo: Nunca sirvas a quem serviu nem peças a quem pediu

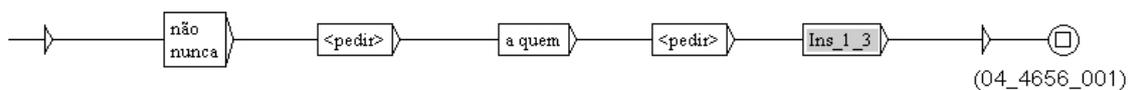


Figura 82: Grafo: Nunca peças a quem pediu nem devas a quem deveu

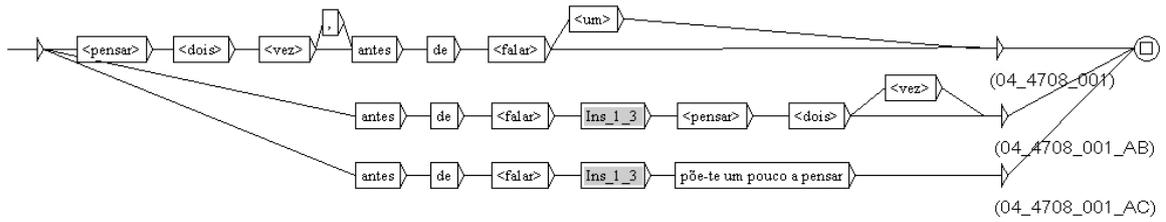


Figura 83: Grafo: *Pensa duas vezes antes de falar*

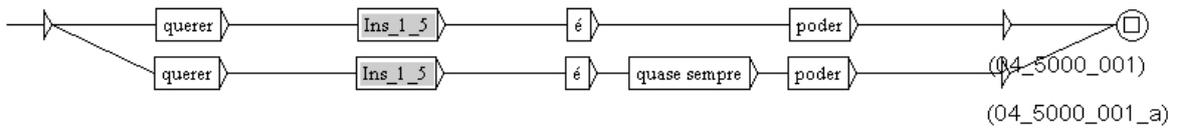


Figura 84: Grafo: *Quereres é poder*

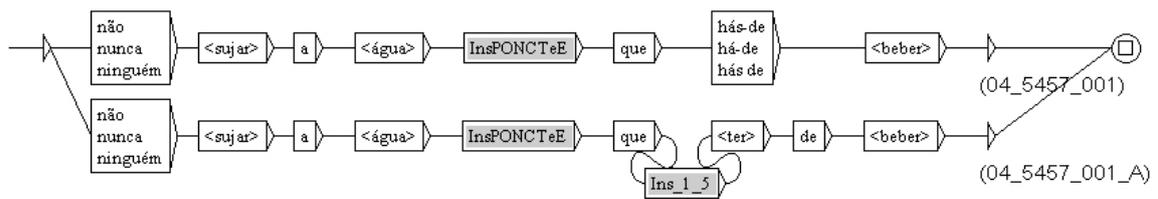


Figura 85: Grafo: *Não sujes a água que há de beber*

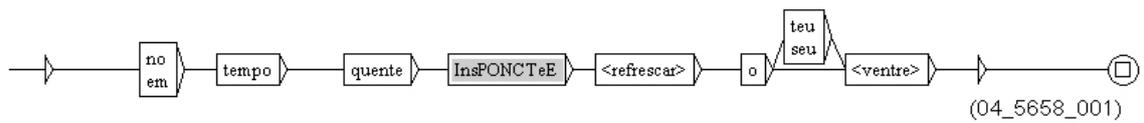


Figura 86: Grafo: *No tempo quente refresca o ventre*

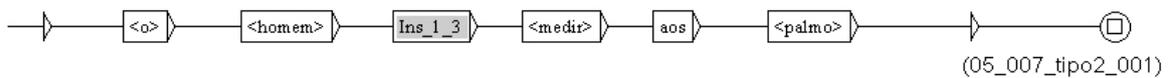


Figura 87: Grafo: *Os homens não se medem aos palmos*

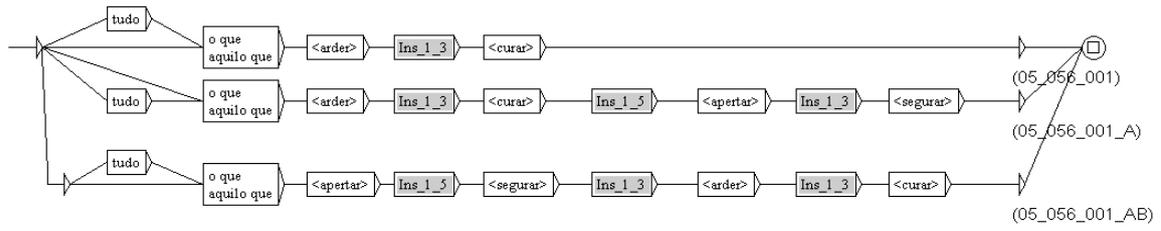


Figura 88: Grafo: O que arde cura

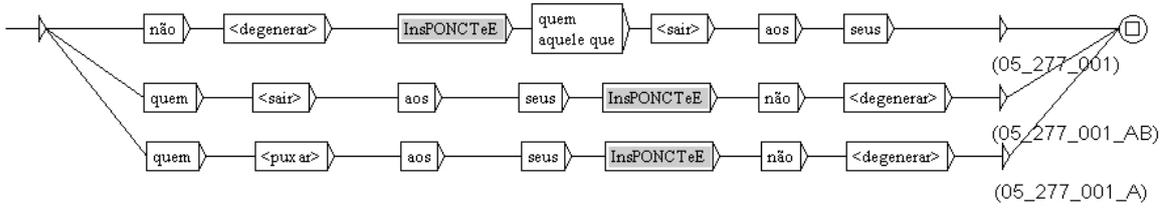


Figura 89: Grafo: Quem sai aos seus não degenera

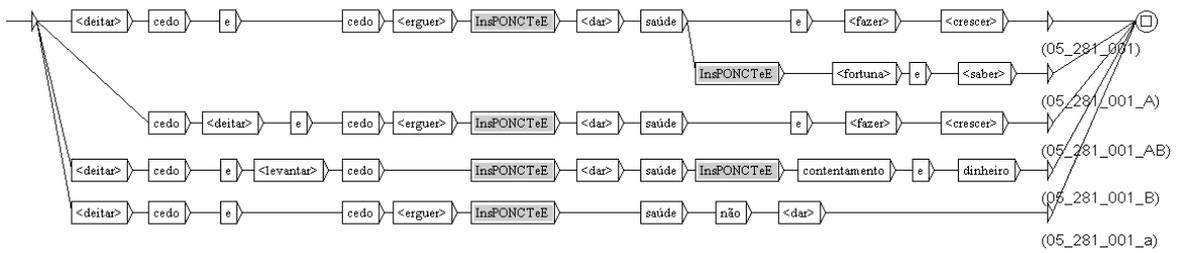


Figura 90: Grafo: Deitar cedo e cedo erguer dá saúde e faz crescer

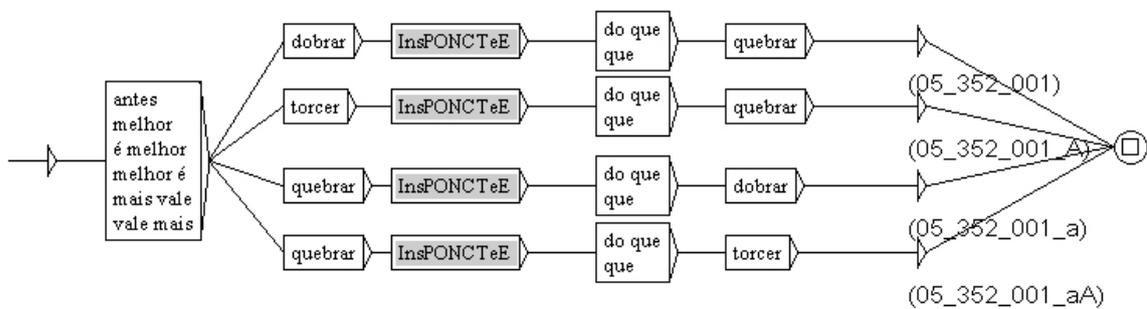


Figura 91: Grafo: Antes dobrar que quebrar

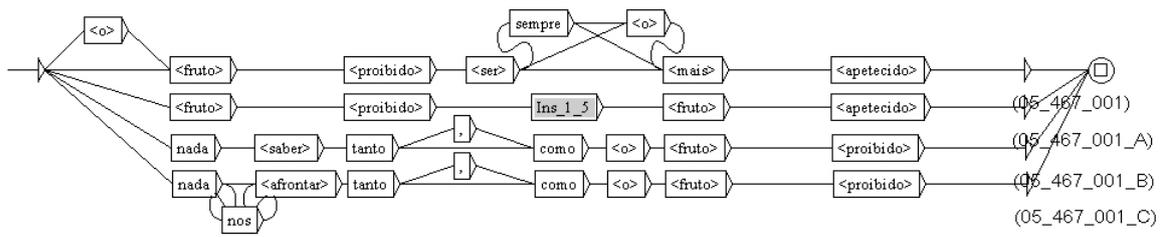


Figura 92: Grafo: O fruto proibido é sempre o mais apetecido

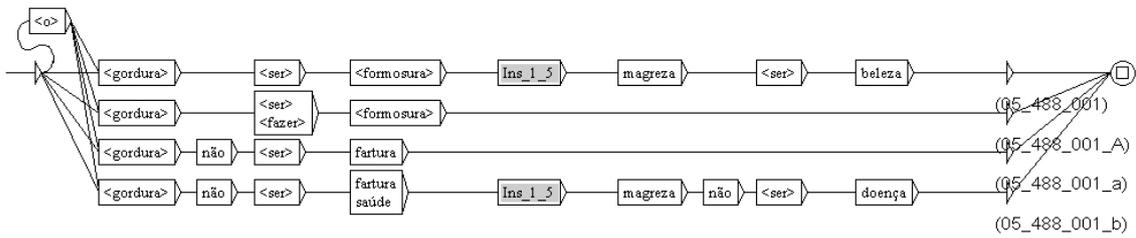


Figura 93: Grafo: Gordura é formosura

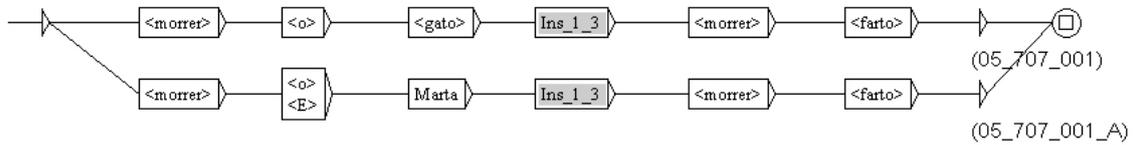


Figura 94: Grafo: Morra Marta morra farta

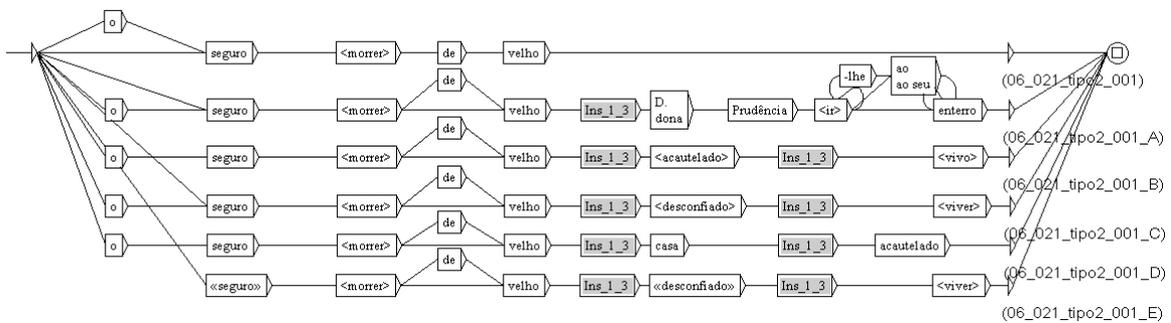


Figura 95: Grafo: O seguro morreu de velho

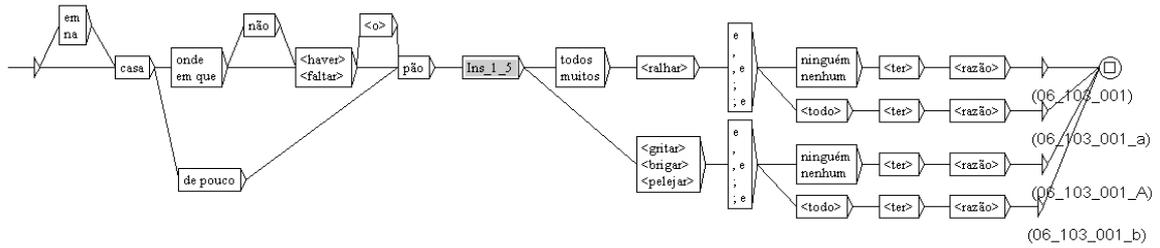


Figura 96: Grafo: Casa onde não há pão todos ralham e ninguém tem razão

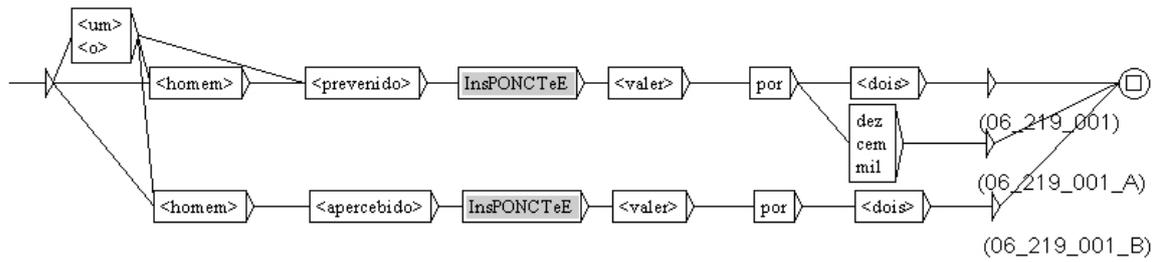


Figura 97: Grafo: Um homem prevenido vale por dois

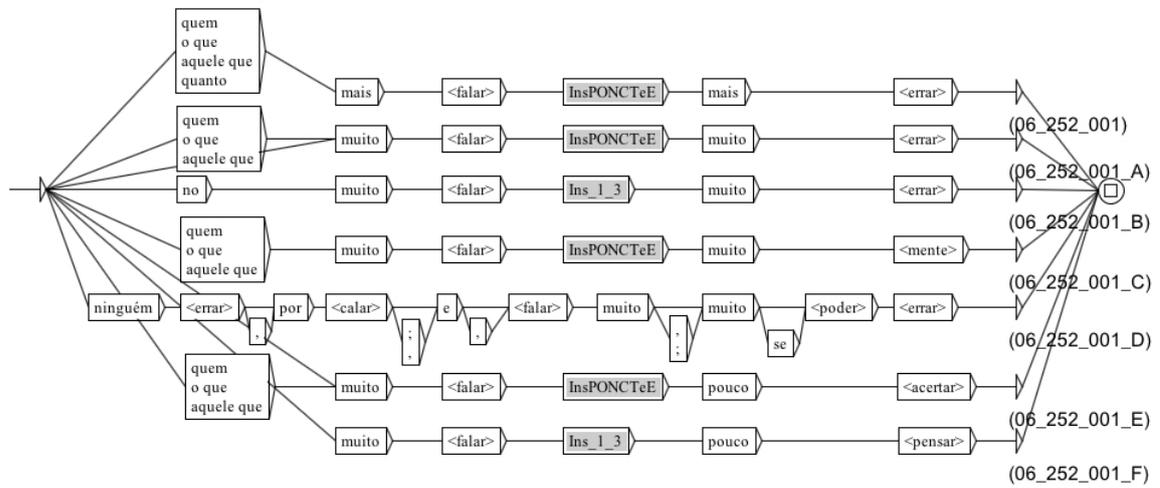


Figura 98: Grafo: Quem mais fala mais erra

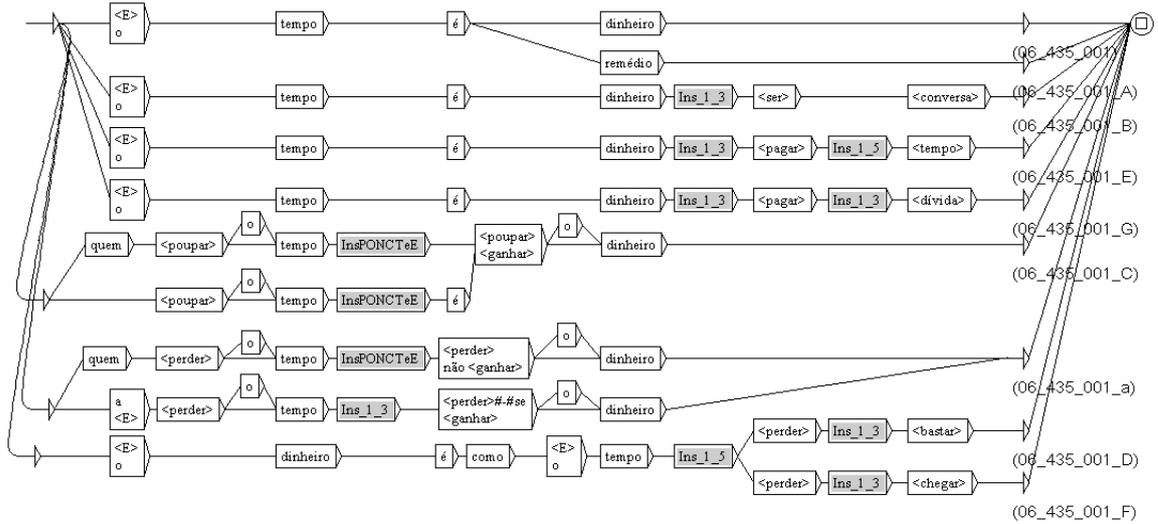


Figura 99: Grafo: Tempo é dinheiro

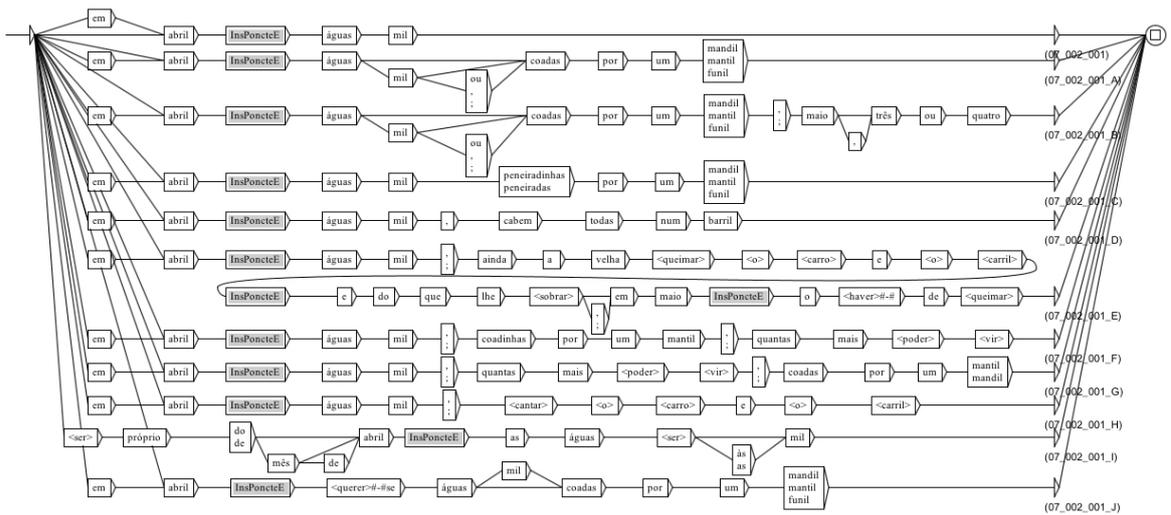


Figura 100: Grafo: Abril águas mil

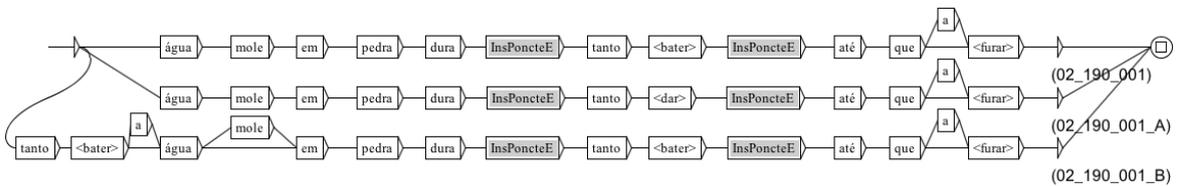


Figura 101: Grafo: Água mole em pedra dura tanto dá até que fura

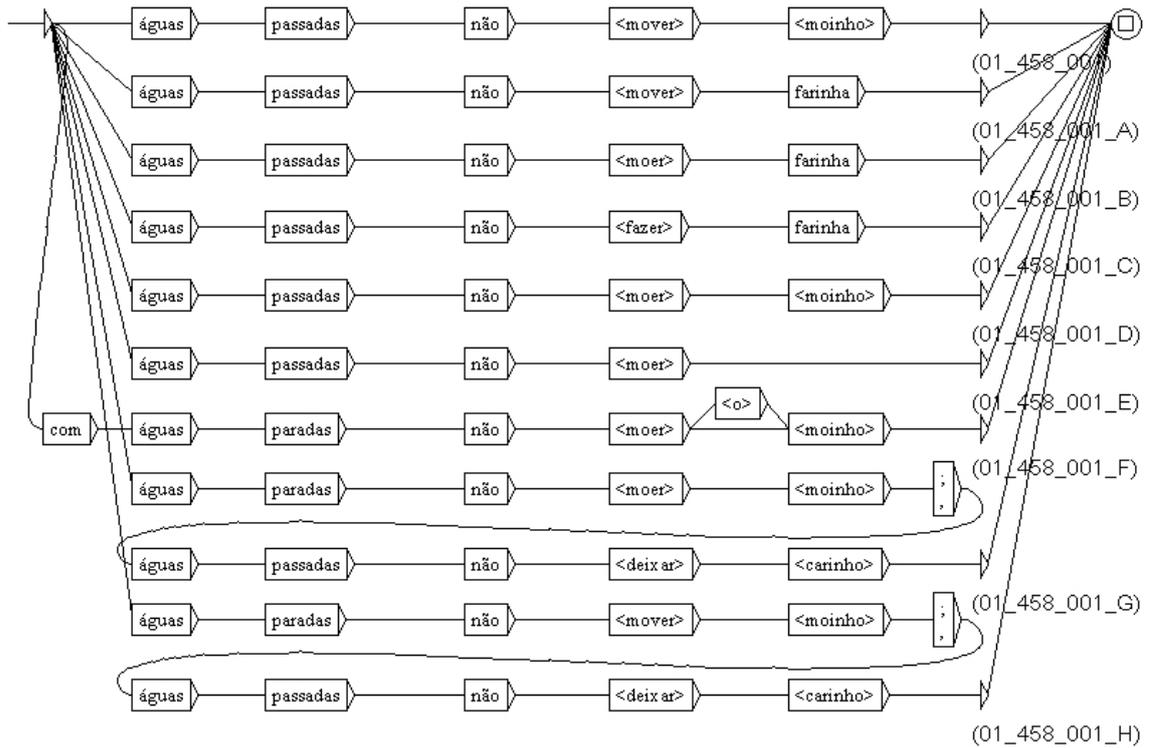


Figura 102: Grafo: Águas passadas não movem moinhos

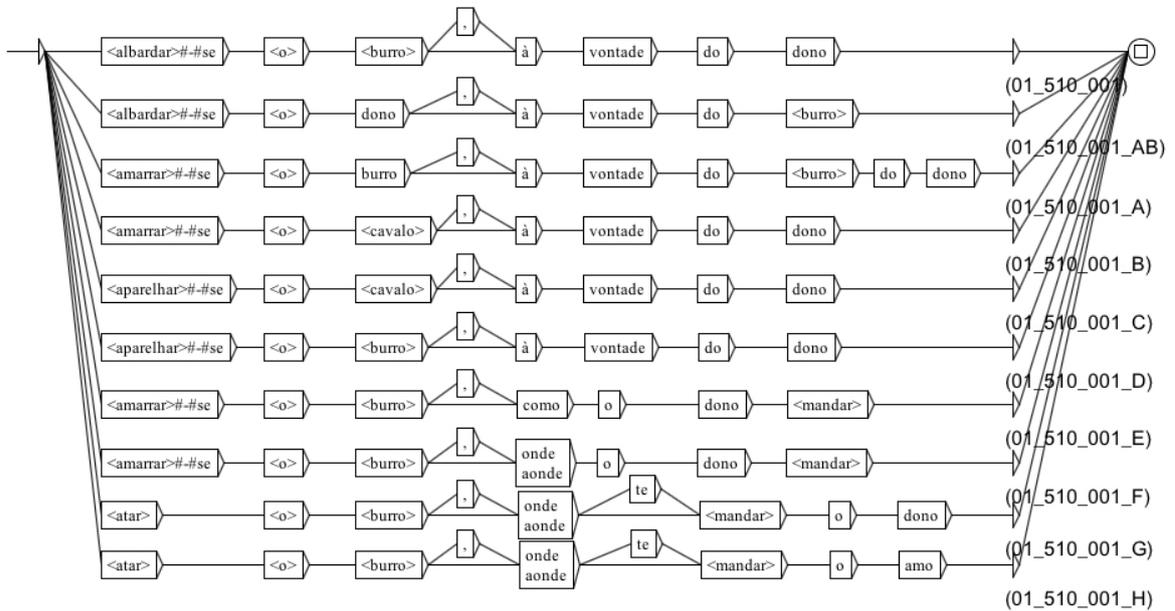


Figura 103: Grafo: Albarda-se o burro à vontade do dono

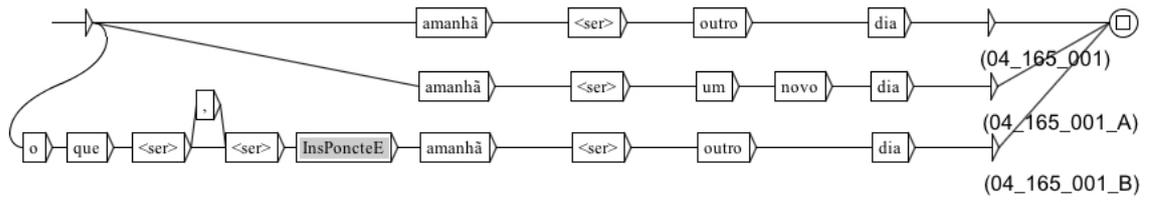


Figura 104: Grafo: *Amanhã é outro dia*



Figura 105: Grafo: *Amanhã também é dia*

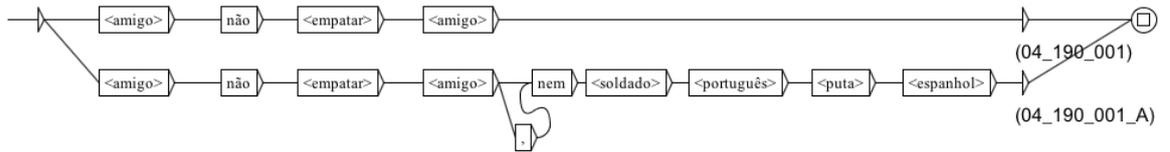


Figura 106: Grafo: *Amigo não empata amigo*



Figura 107: Grafo: *Amigos, amigos, negócios à parte*

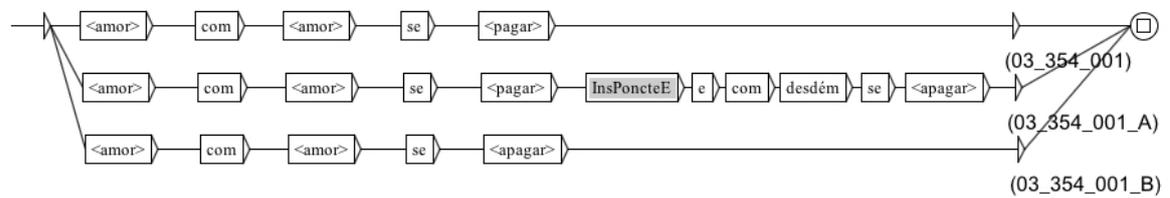


Figura 108: Grafo: *Amor com amor se paga*

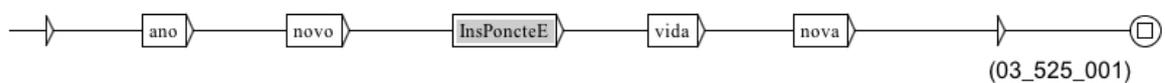


Figura 109: Grafo: *Ano novo, vida nova*

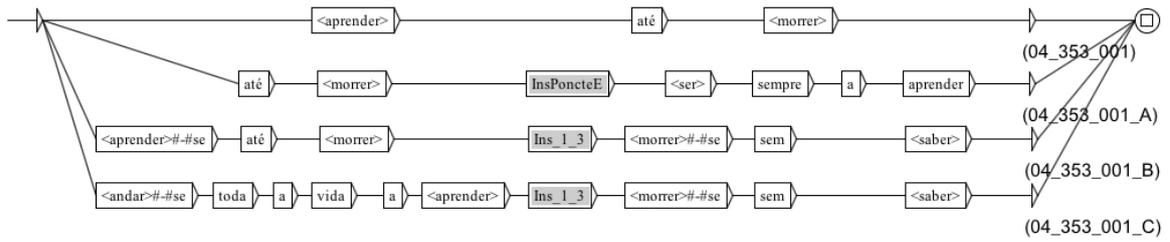


Figura 110: Grafo: Aprender até morrer

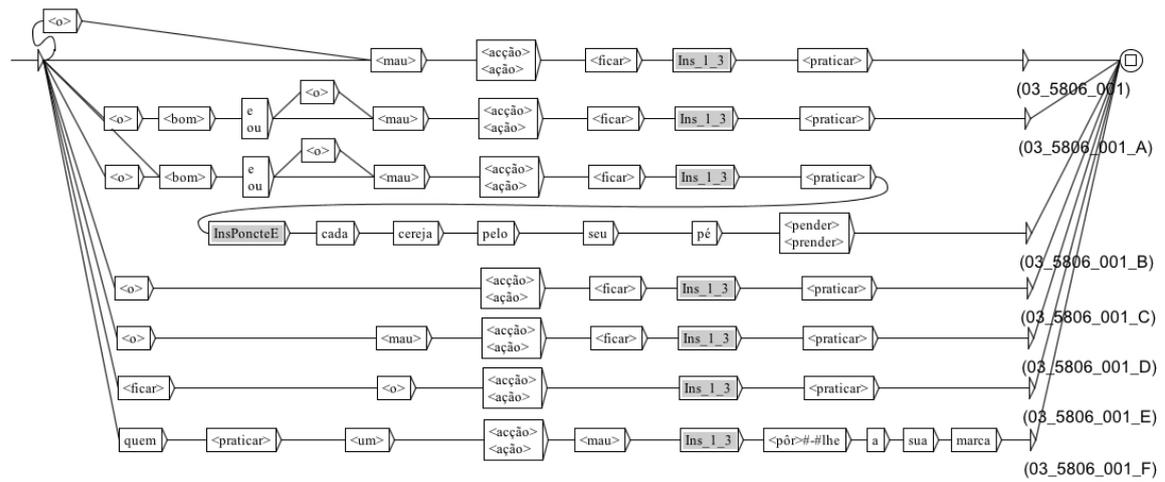


Figura 111: Grafo: A má ação fica com quem a pratica

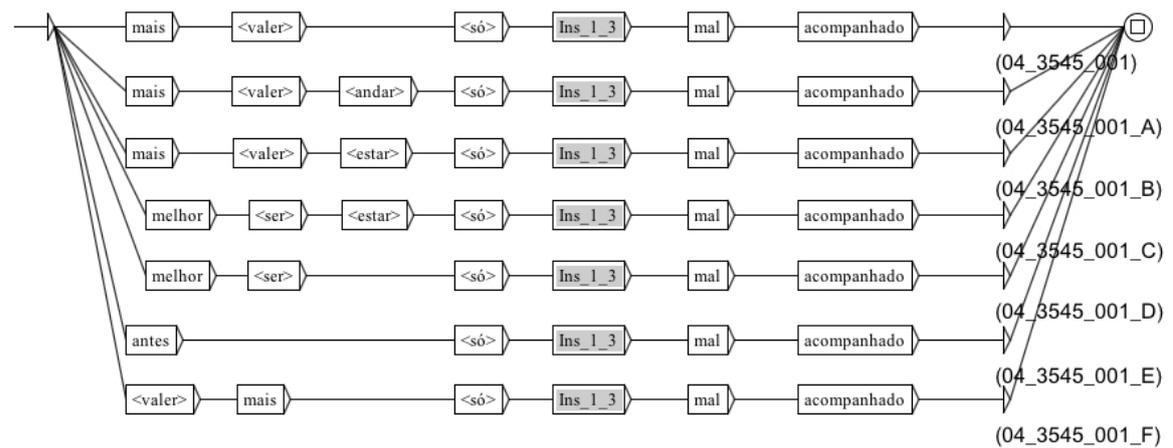


Figura 112: Grafo: Mais vale só que mal acompanhado

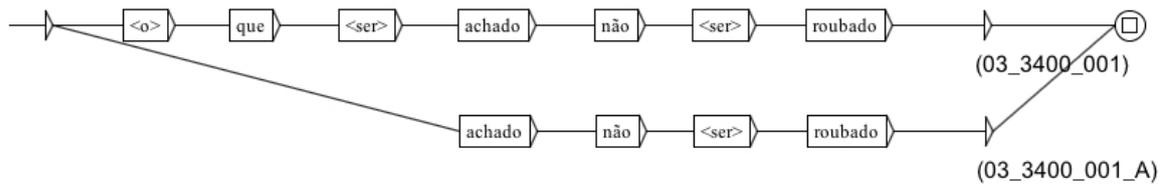


Figura 113: Grafo: O que é achado não é roubado

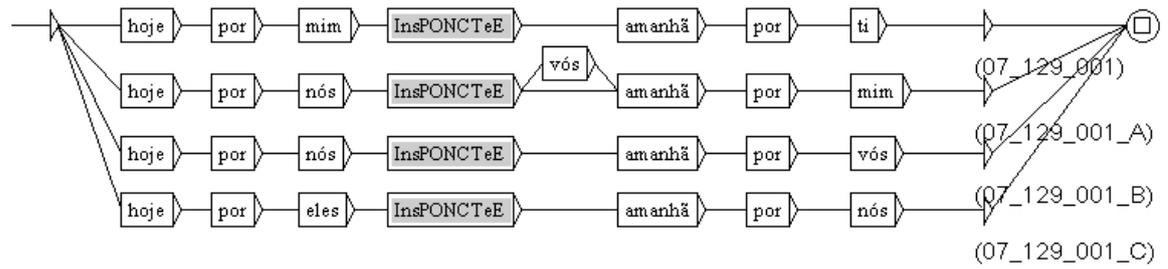


Figura 114: Grafo: Hoje por mim amanhã por ti

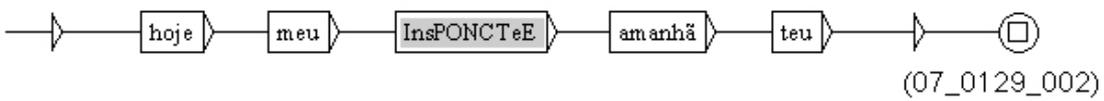


Figura 115: Grafo: Hoje meu amanhã teu

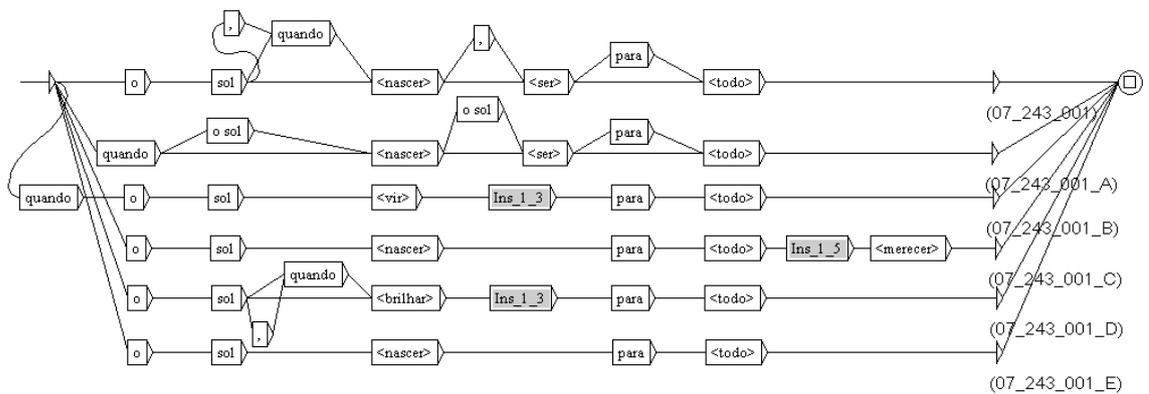


Figura 116: Grafo: O sol quando nasce é para todos

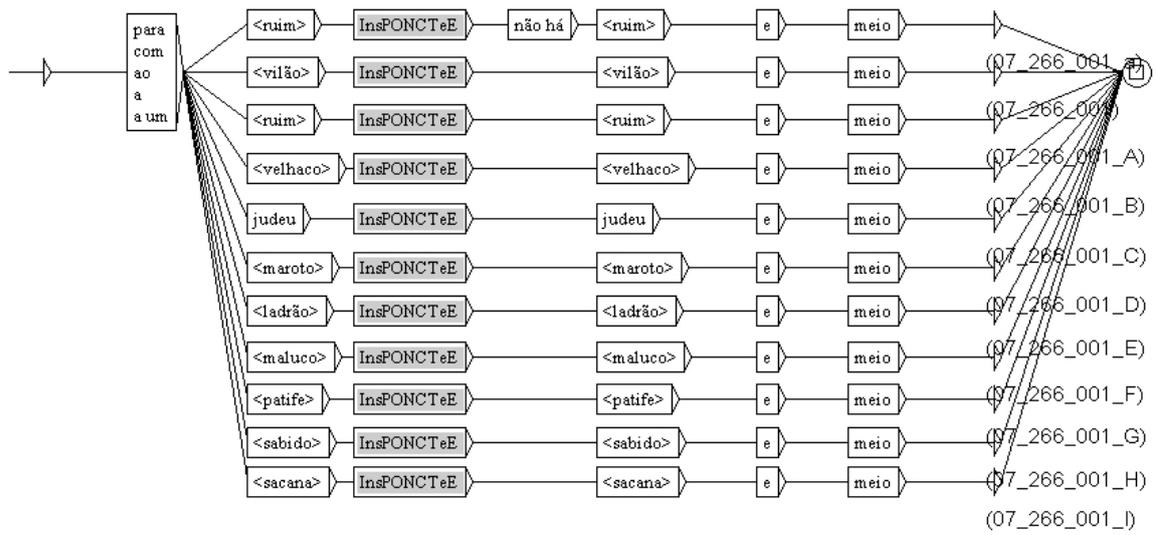


Figura 117: Grafo: Para vilão, vilão e meio

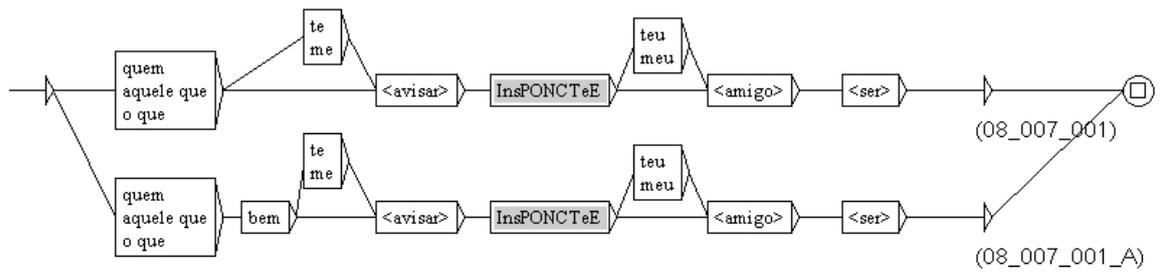


Figura 118: Grafo: Quem te avisa teu amigo é

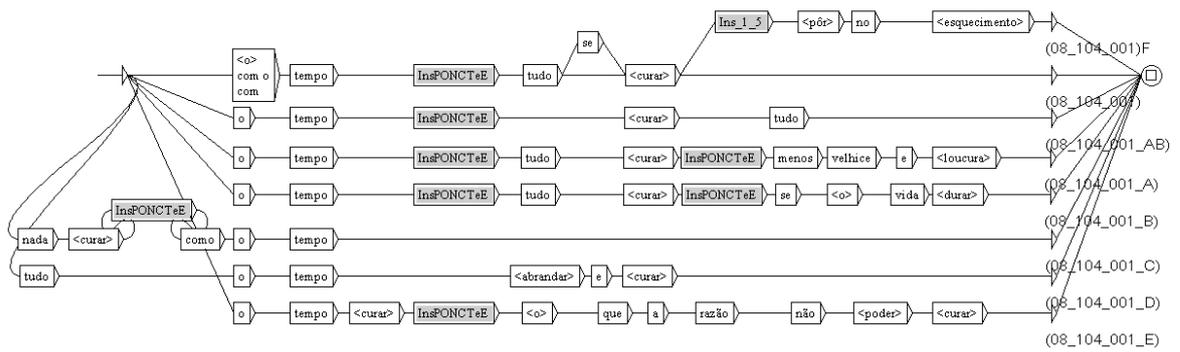


Figura 119: Grafo: Com o tempo tudo se cura

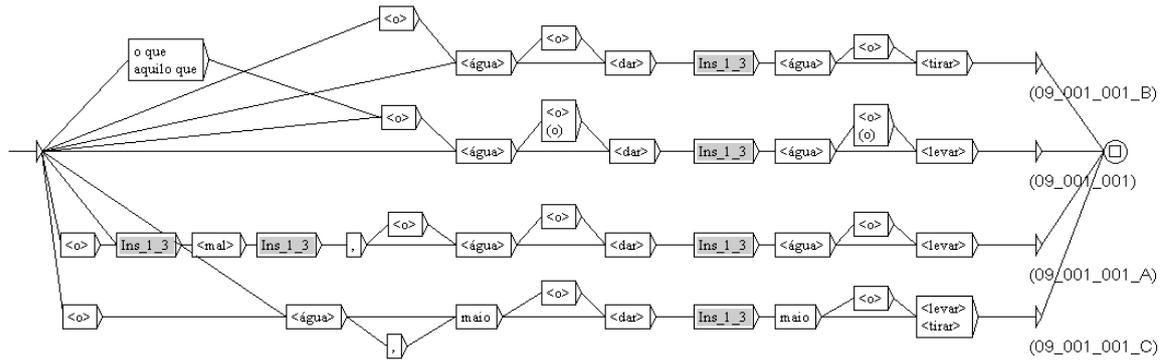


Figura 120: Grafo: A água o dá a água o leva

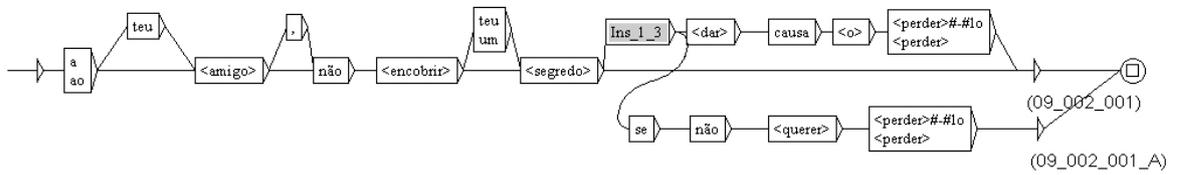


Figura 121: Grafo: Ao teu amigo não encubras segredo

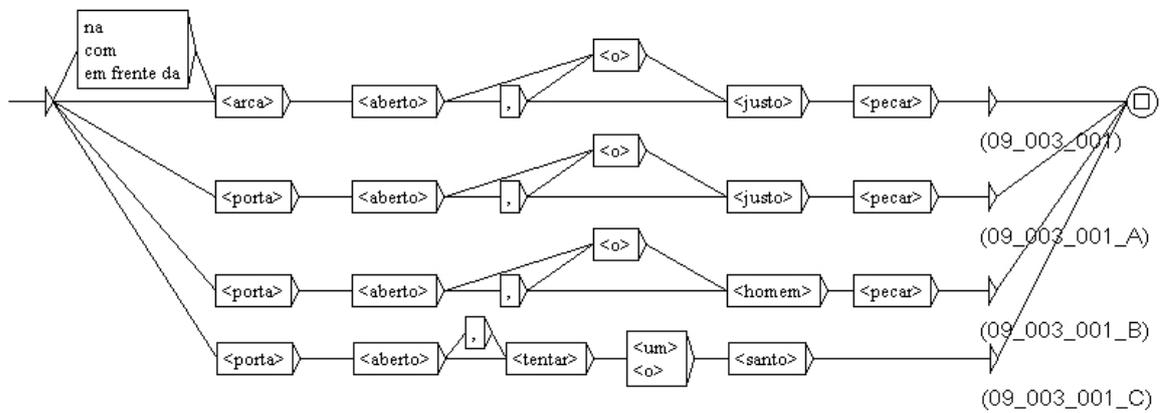


Figura 122: Grafo: Na arca aberta o justo peca

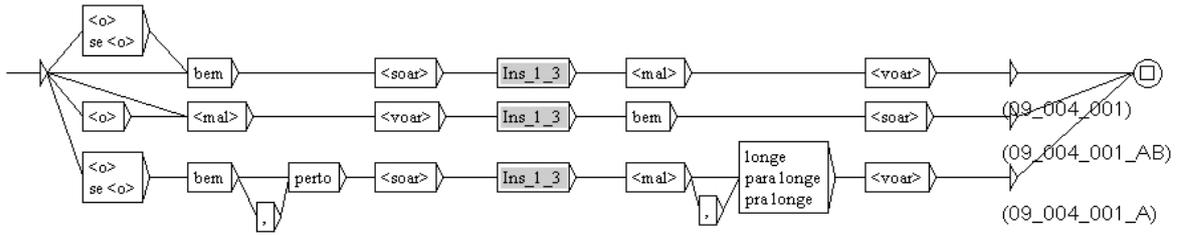


Figura 123: Grafo: O bem soa o mal voa

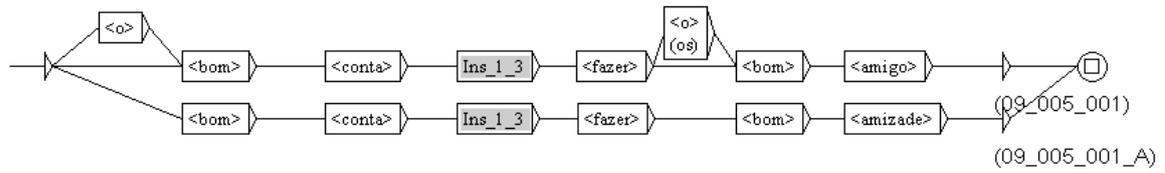


Figura 124: Grafo: Boas contas fazem bons amigos

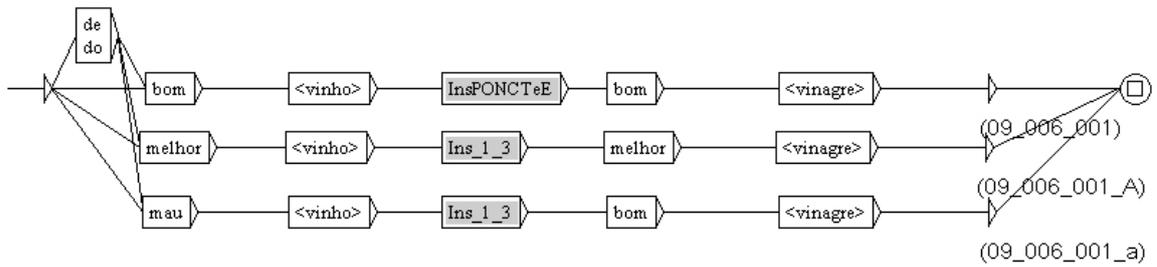


Figura 125: Grafo: Bom vinho bom vinagre

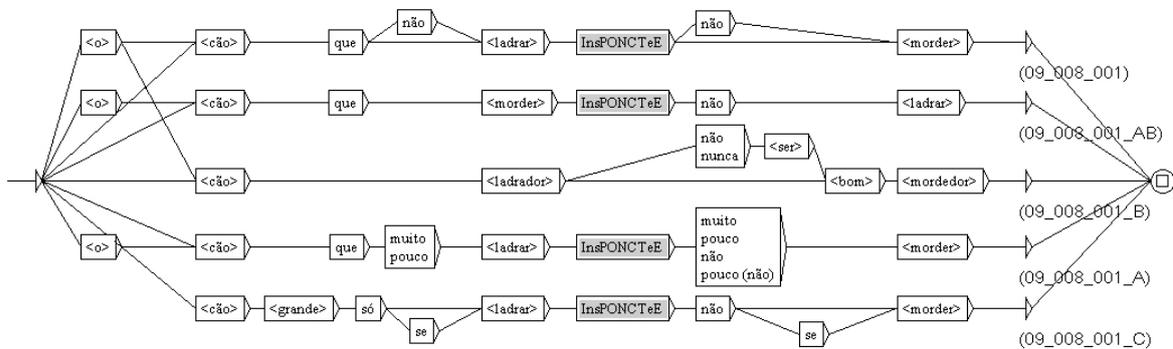


Figura 126: Grafo: Cão que ladra não morde

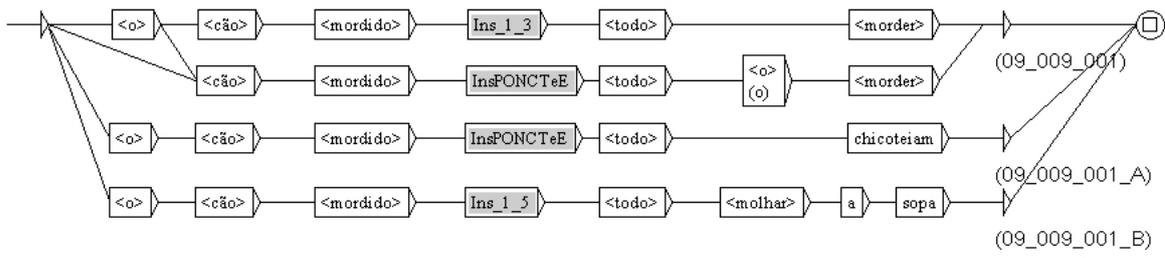


Figura 127: Grafo: A cão mordido todos o mordem

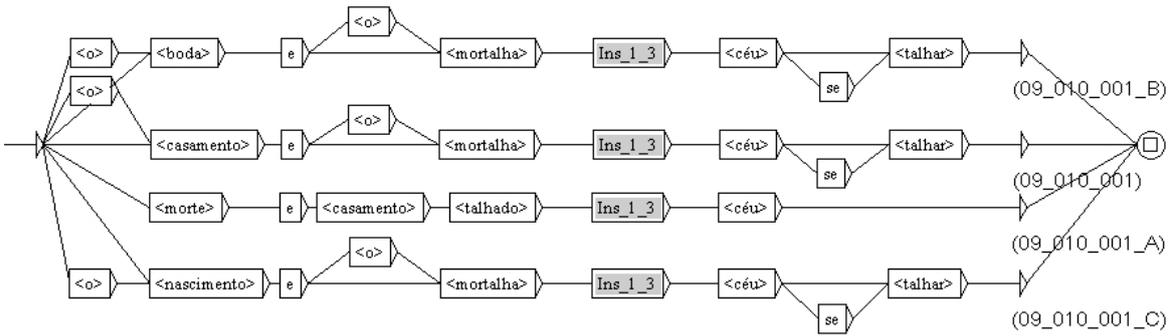


Figura 128: Grafo: A boda e a mortalha no céu se talha

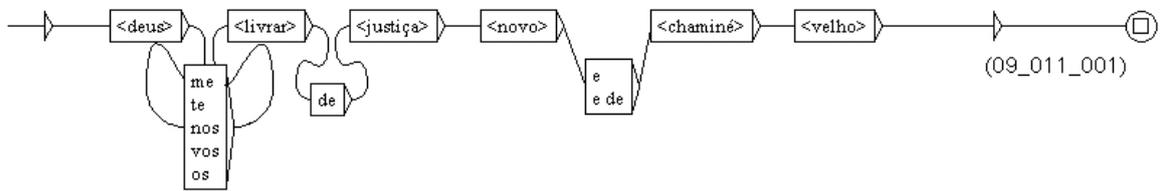


Figura 129: Grafo: Deus nos livre de justiça novas e de chaminés velhas

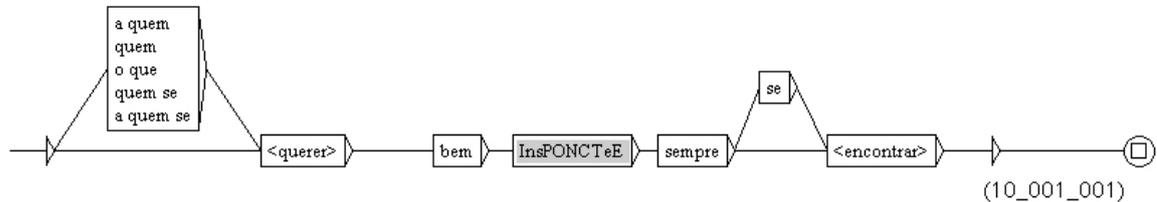


Figura 130: Grafo: Quem se quer bem sempre se encontra

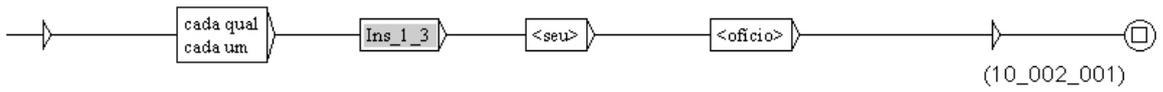


Figura 131: Grafo: Cada um com o seu ofício

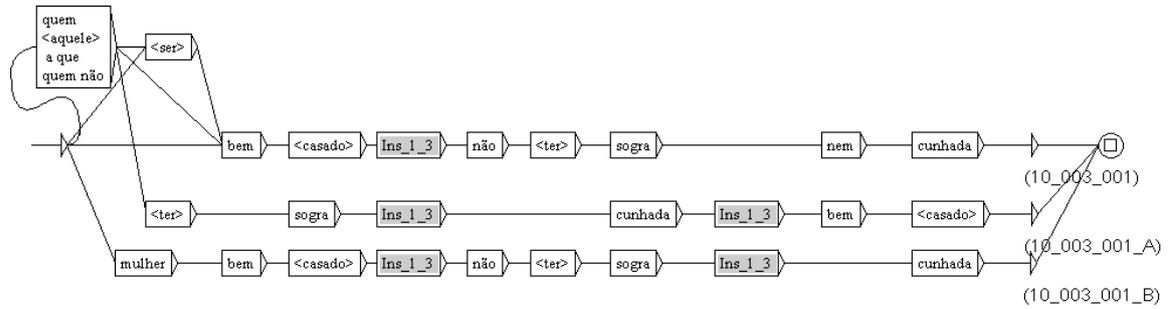


Figura 132: Grafo: É bem casada a que não tem sogra nem cunhada

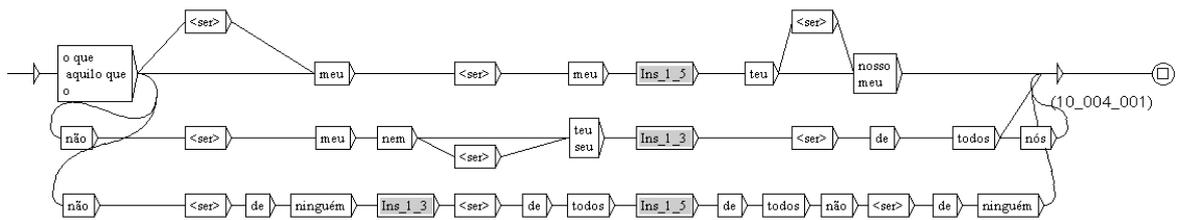


Figura 133: Grafo: O que é meu é meu o que é teu é nosso

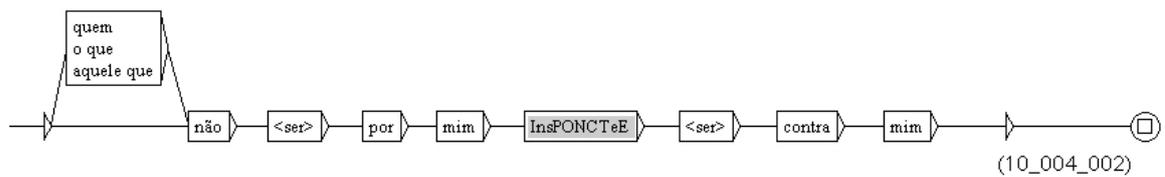


Figura 134: Grafo: Quem não é por mim é contra mim

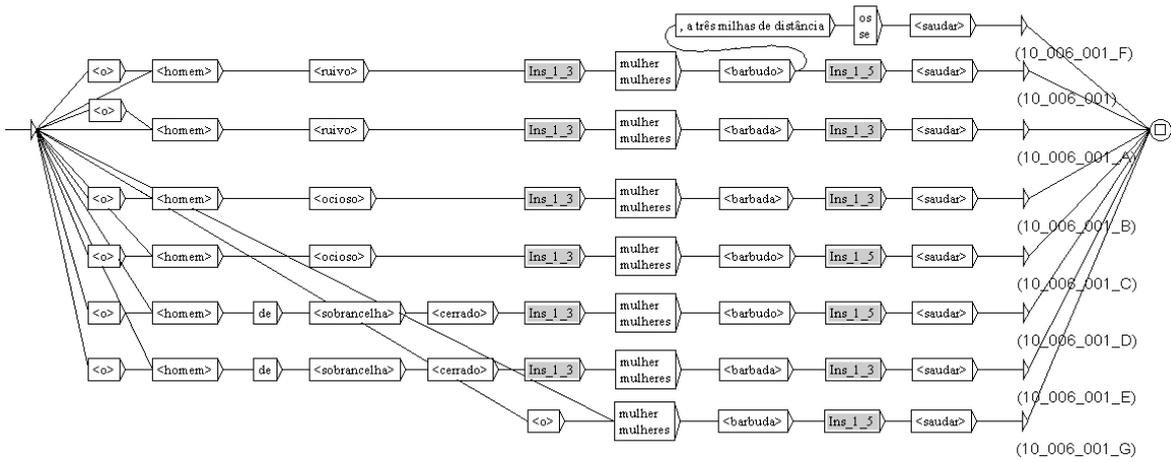


Figura 135: Grafo: *A homem ruivo e mulher barbuda de longe os saúda*

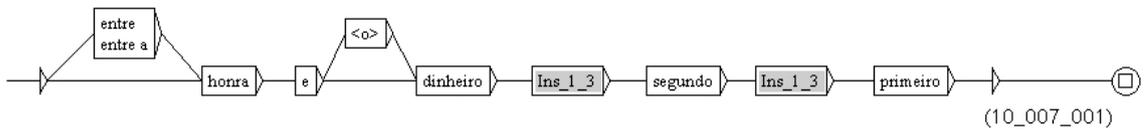


Figura 136: Grafo: *Entre a honra e o dinheiro o segundo é o primeiro*

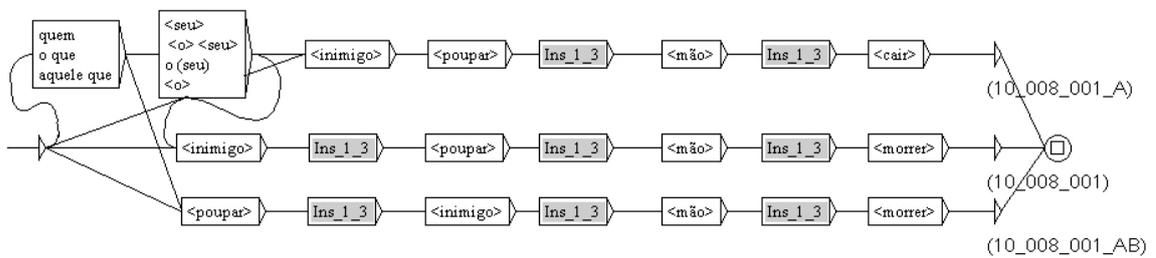


Figura 137: Grafo: *Quem o seu inimigo poupa nas mãos lhe morre*

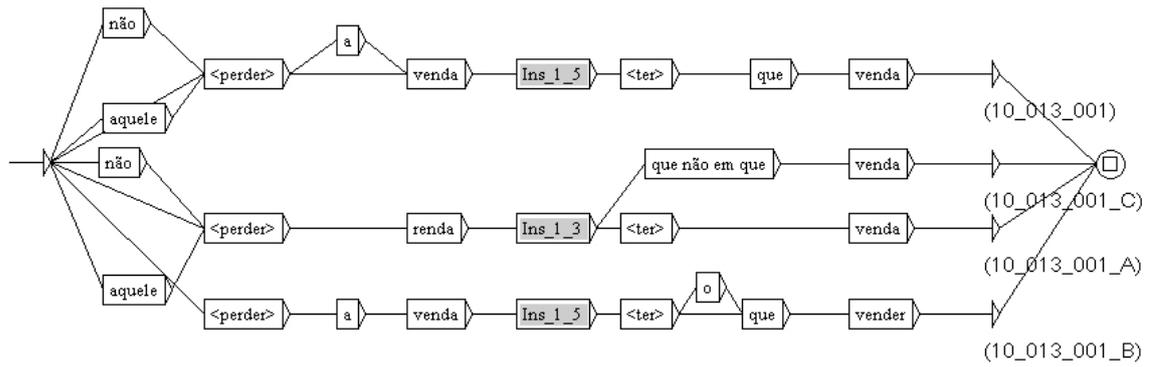


Figura 142: Grafo: Não perde venda senão quem não tem que venda

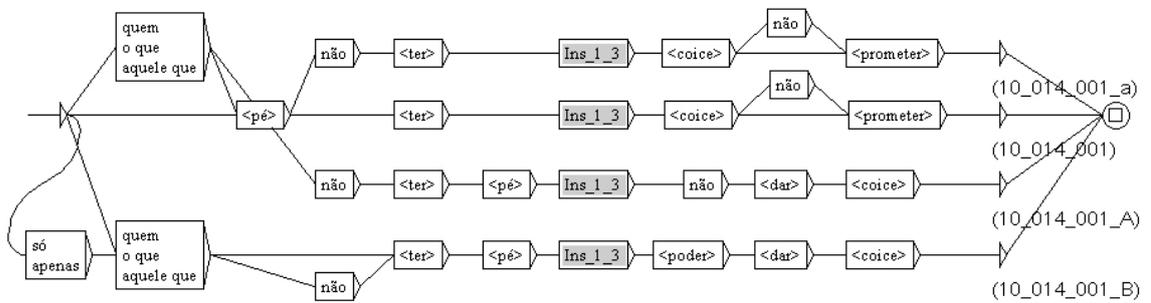


Figura 143: Grafo: Quem pés não tem coices não promete

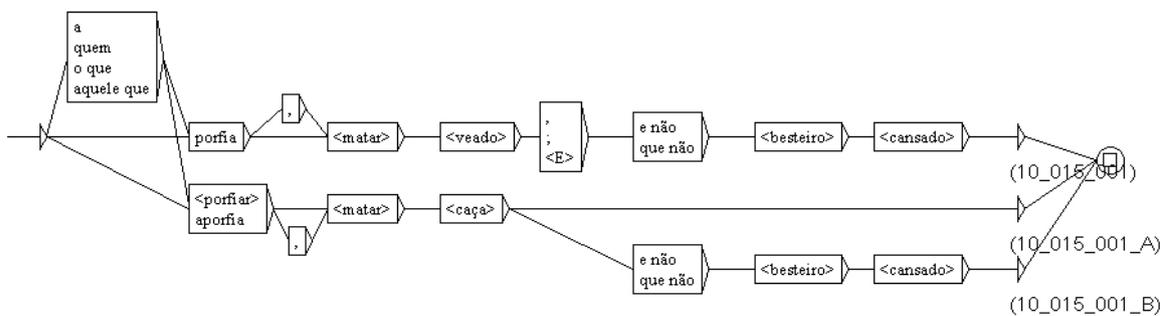


Figura 144: Grafo: Porfia mata veado e não besteiro cansado

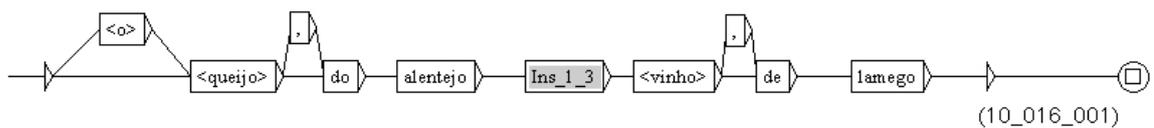


Figura 145: Grafo: Queijo do Alentejo, vinho de Lamego

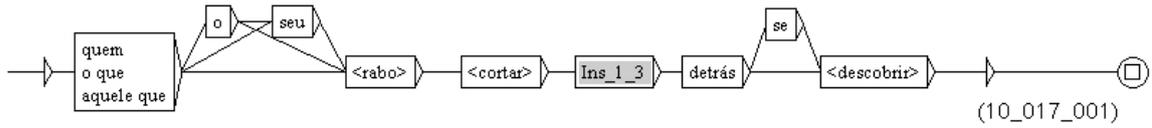


Figura 146: Grafo: Quem o rabo corta por trás se descobre

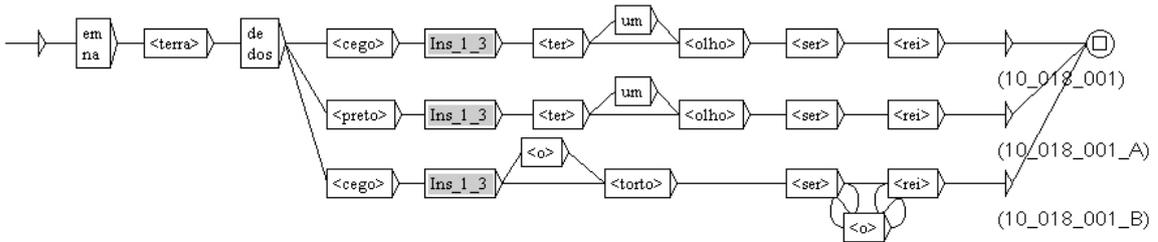


Figura 147: Grafo: Em terra de cegos quem tem um olho é rei

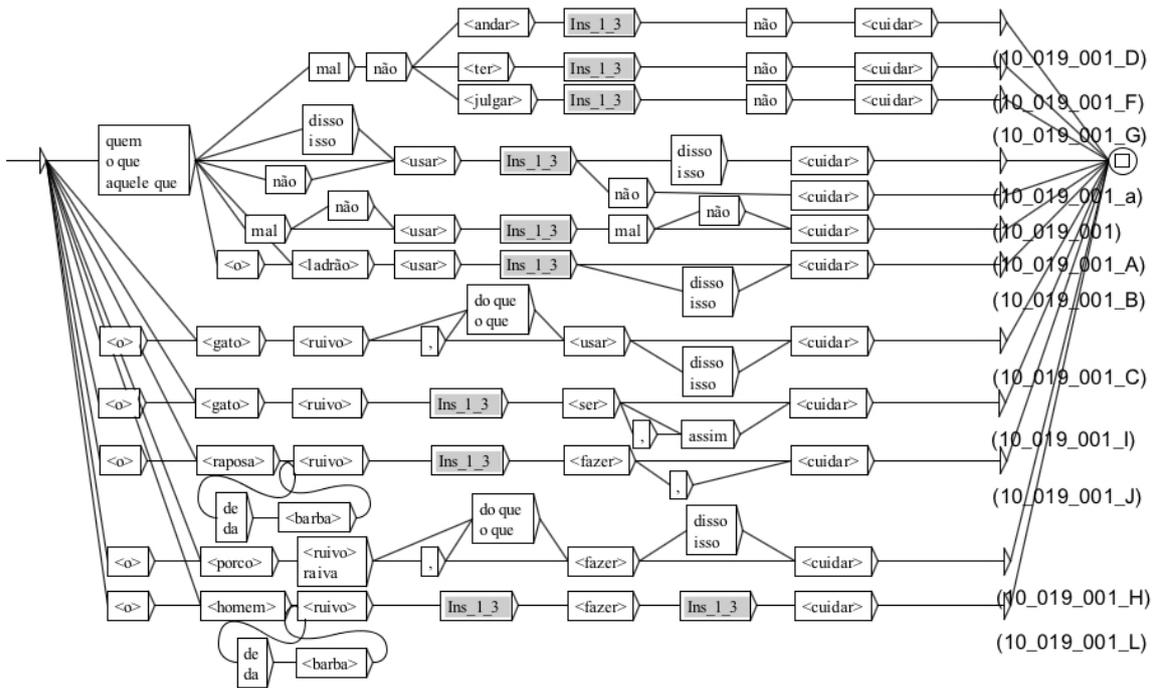


Figura 148: Grafo: Quem não usa não cuida

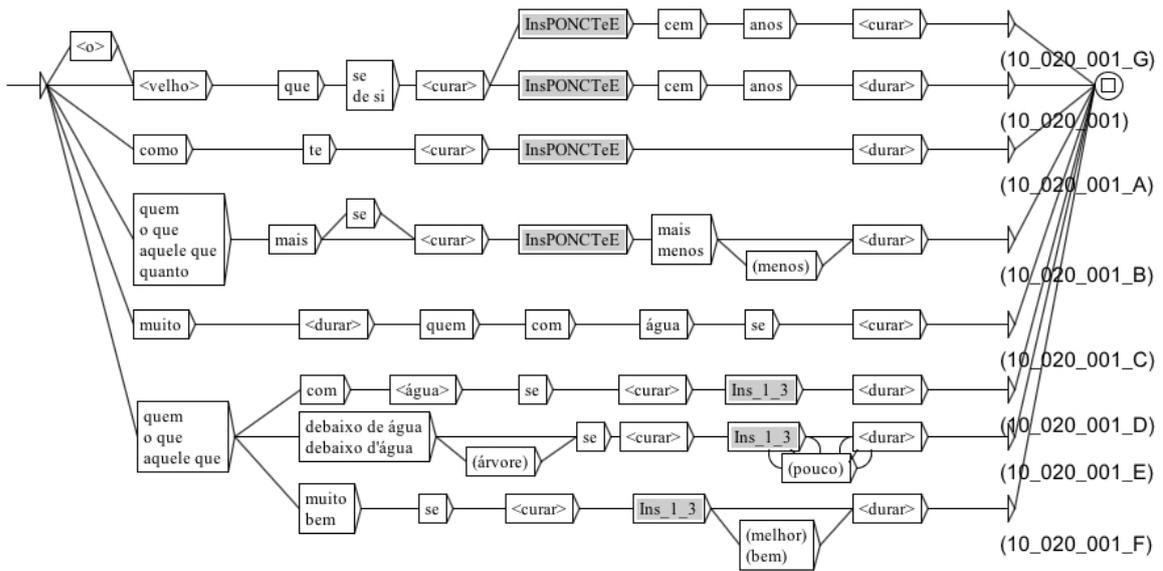


Figura 149: Grafo: Velho que se si cuida cem anos dura

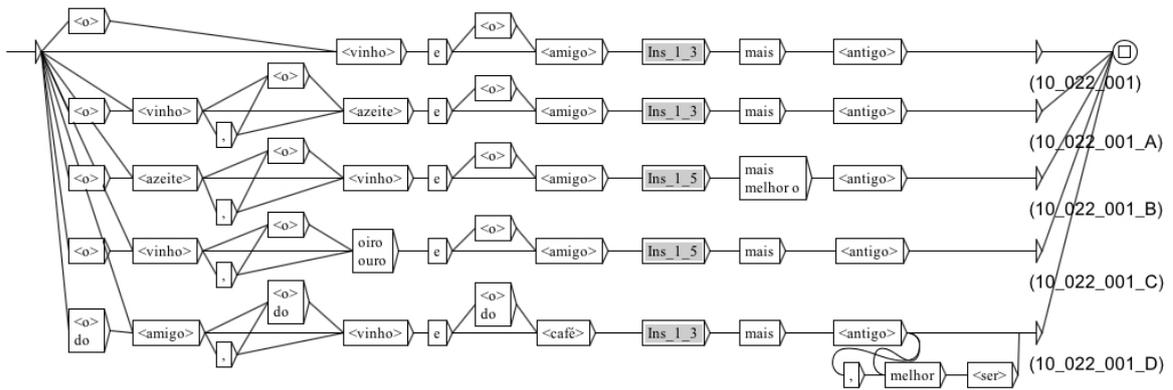


Figura 150: Grafo: O vinho e o amigo o mais antigo

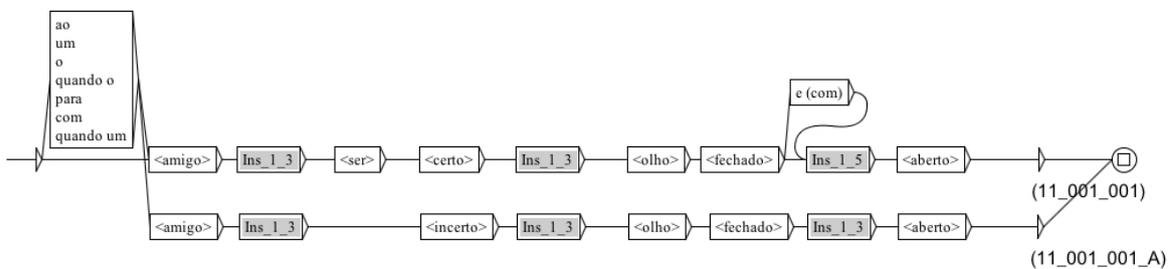


Figura 151: Grafo: Ao amigo que não é certo, um olho fechado e outro aberto

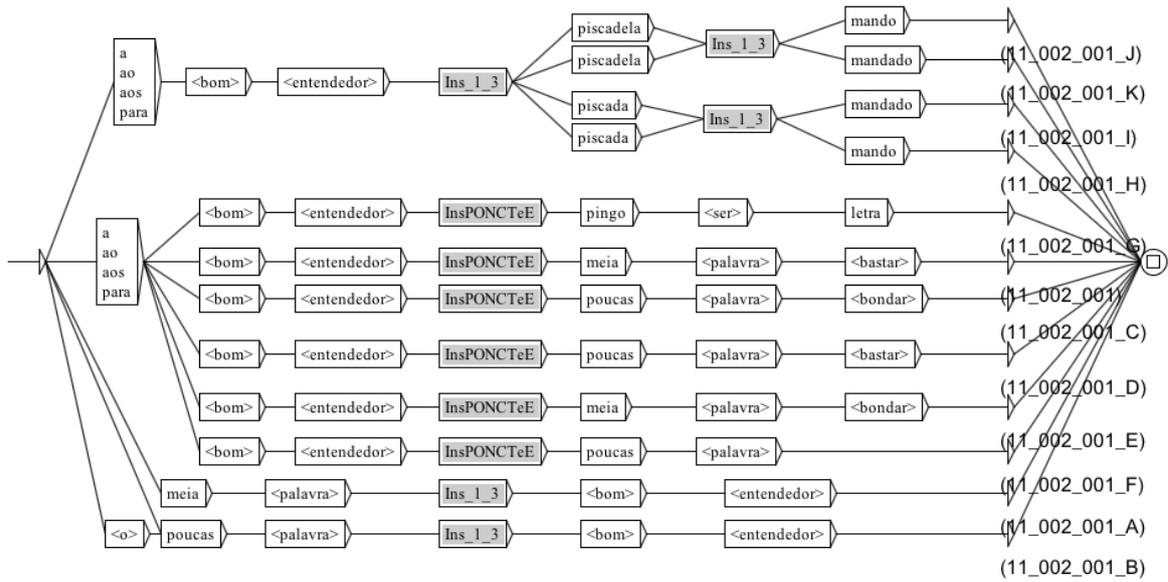


Figura 152: Grafo: Para bom entendedor meia palavra basta

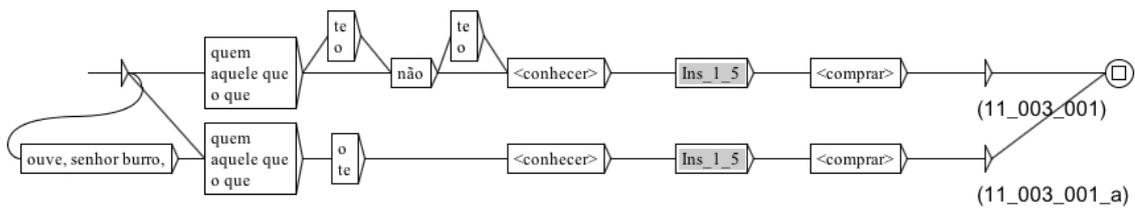


Figura 153: Grafo: Quem não te conhece que te compra

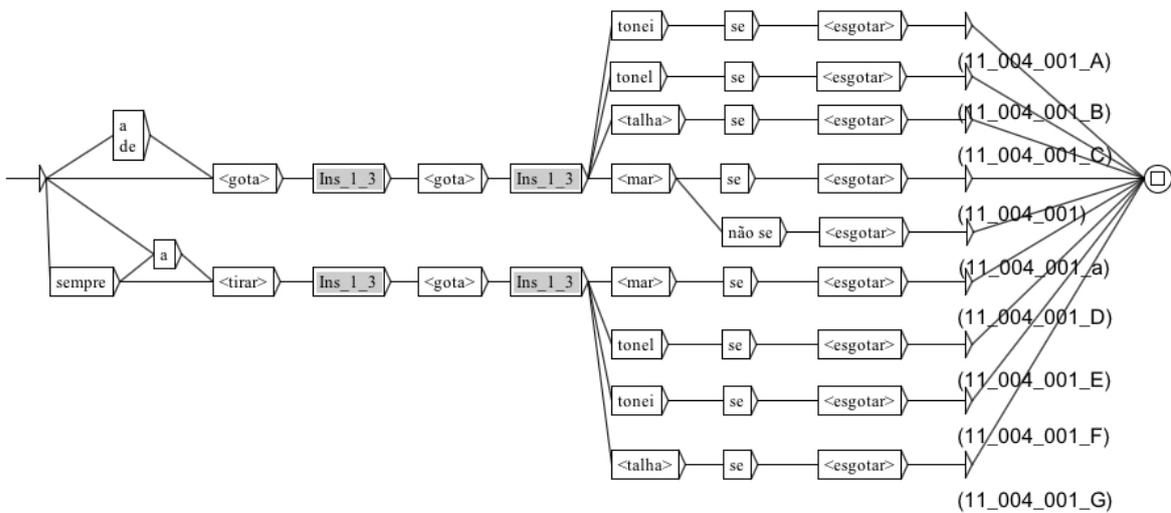


Figura 154: Grafo: Gota a gota o mar se esgota

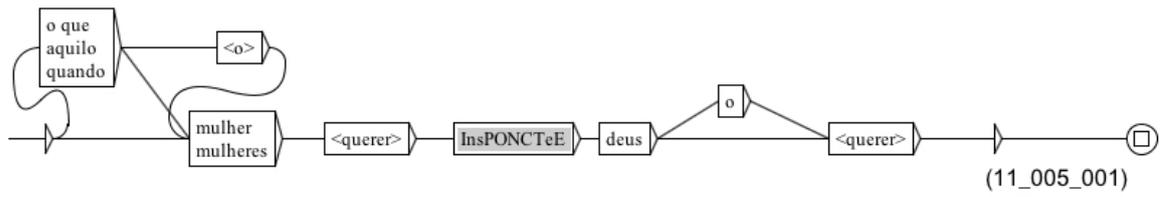


Figura 155: Grafo: *O que a mulher quer Deus quer*

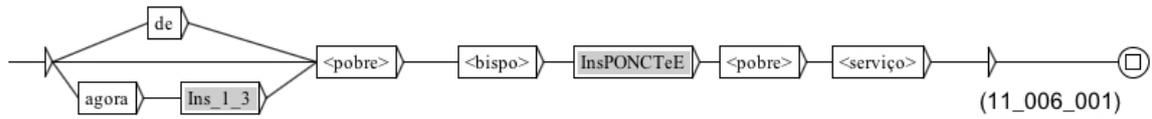


Figura 156: Grafo: *Pobre bispo, pobre serviço*

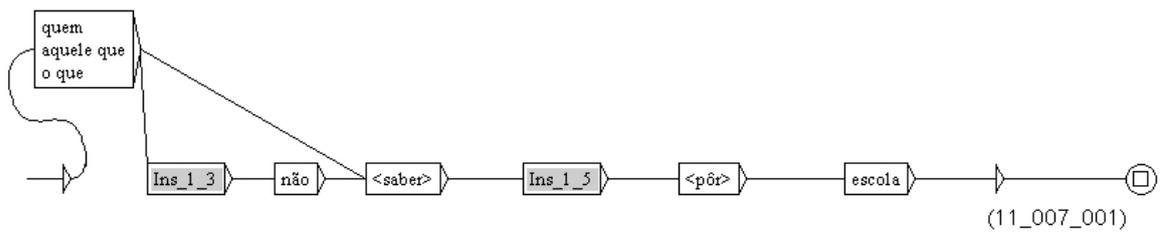


Figura 157: Grafo: *Quem para si não sabe, não ponha escola*

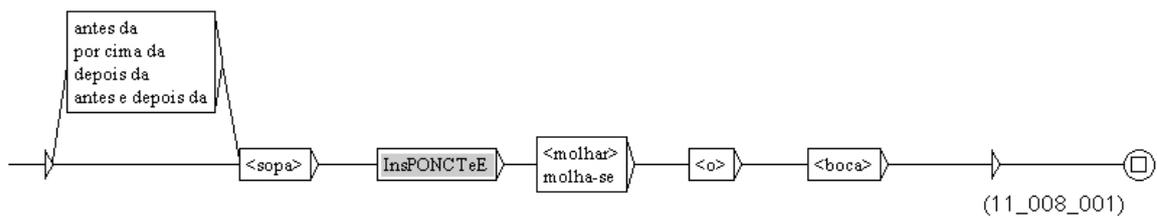


Figura 158: Grafo: *Antes da sopa, molha-se a boca*

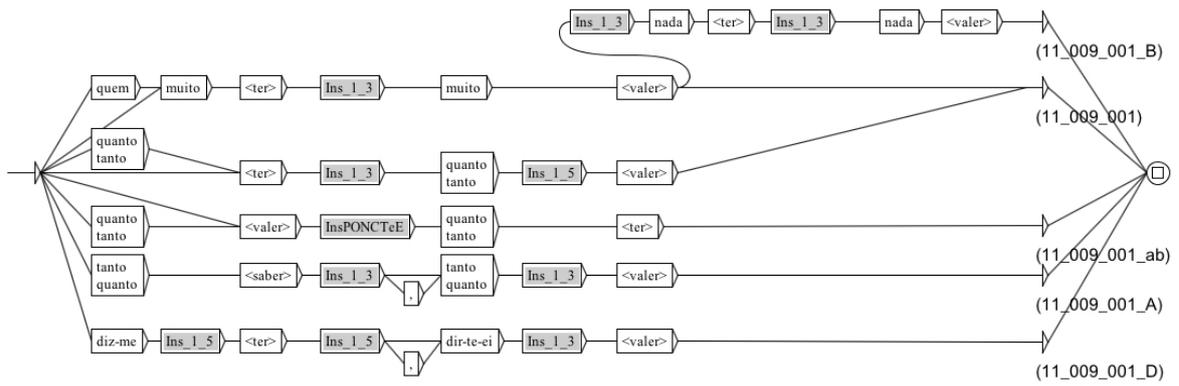


Figura 159: Grafo: Quanto tens, quanto vales

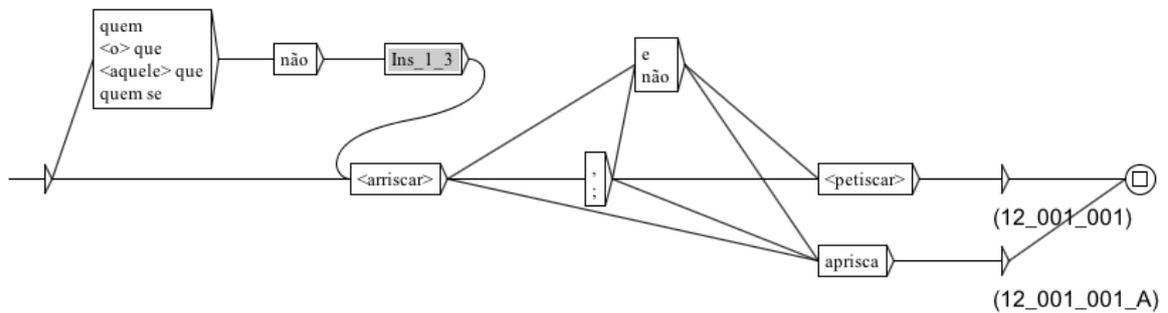


Figura 160: Grafo: Quem não arrisca não petisca

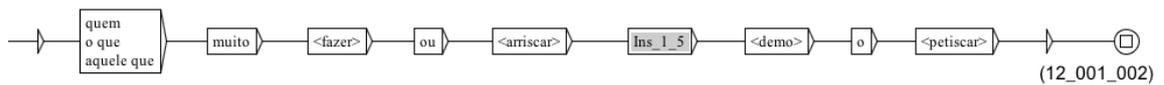


Figura 161: Grafo: Quem muito faz ou arrisca um dia o demo o petisca

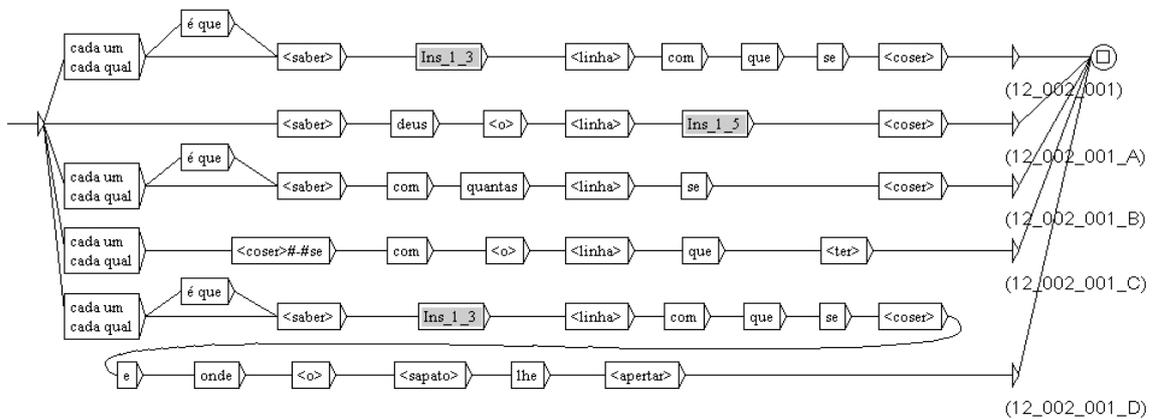


Figura 162: Grafo: Cada um sabe as linhas com que se cose

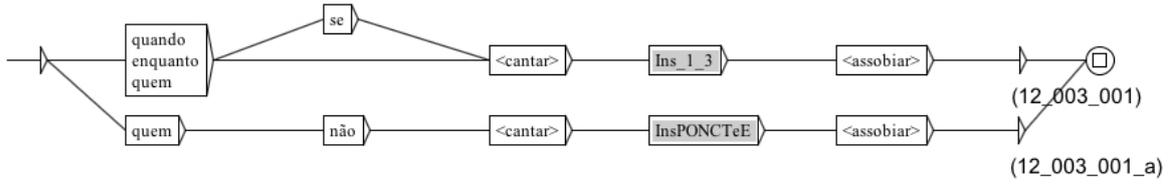


Figura 163: Grafo: *Quem canta não assobia*

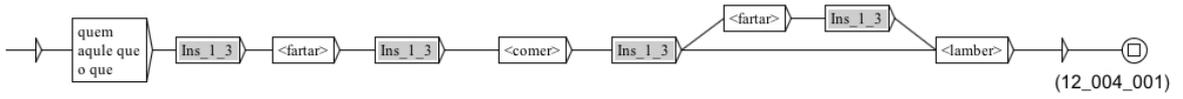


Figura 164: Grafo: *Quem não se farta a comer não se farta a lamber*

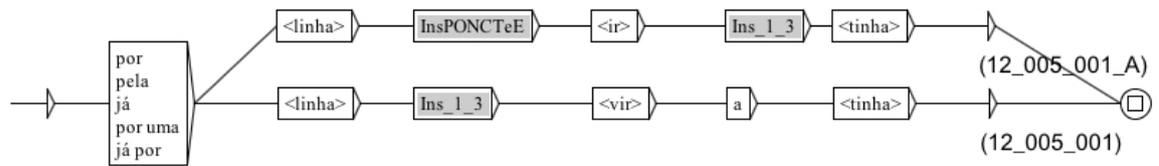


Figura 165: Grafo: *Pela linha vem a linha*

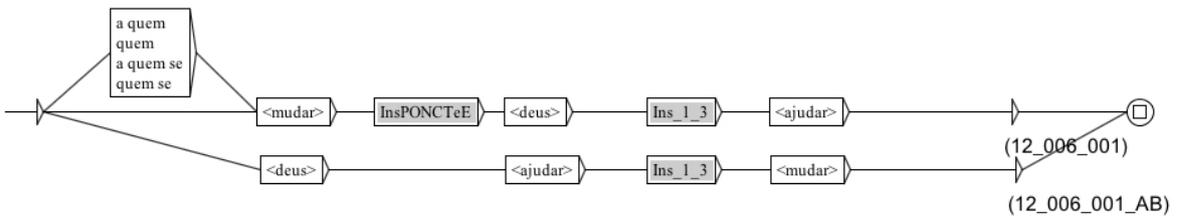


Figura 166: Grafo: *A quem muda Deus o ajuda*

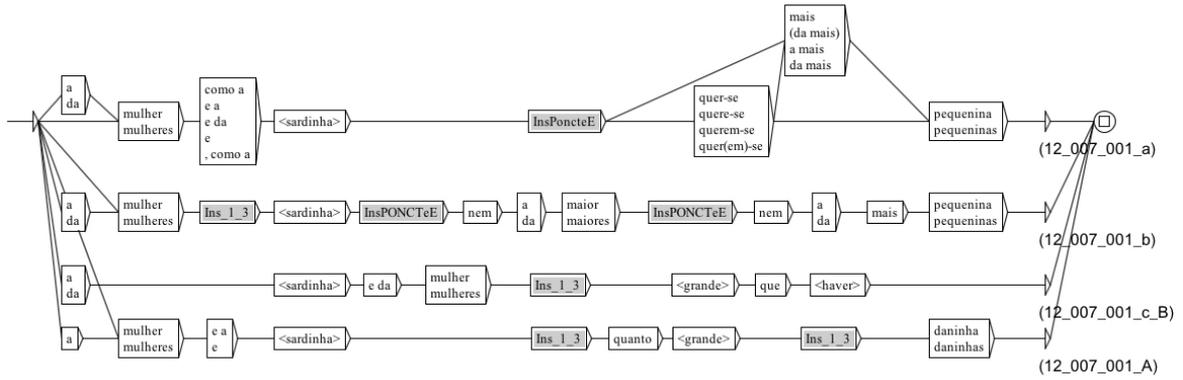


Figura 167: Grafo: A mulher e a sardinha querem-se da mais pequenina

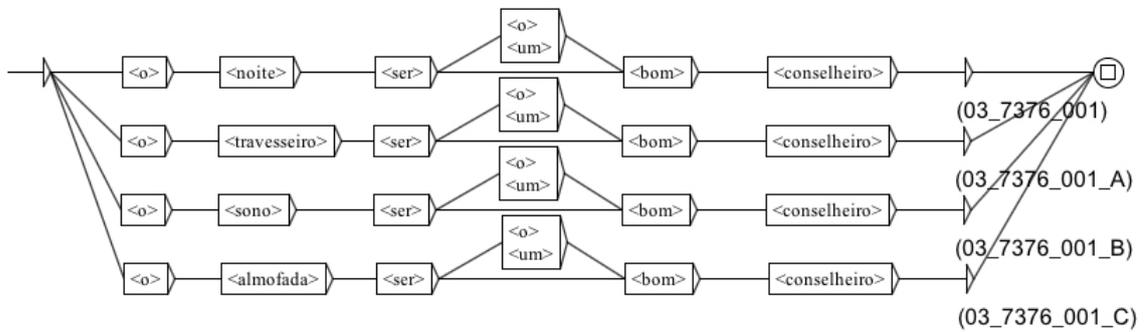


Figura 168: Grafo: A noite é boa conselheira

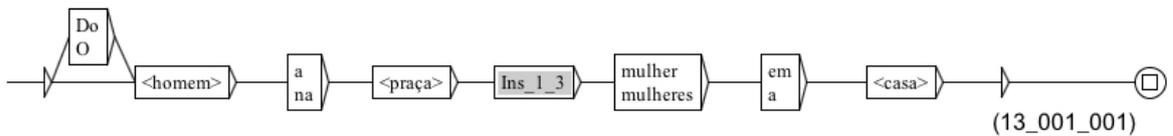


Figura 169: Grafo: O homem na praça e a mulher em casa

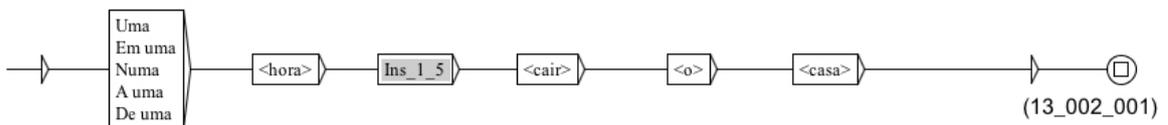


Figura 170: Grafo: A uma hora cai a casa

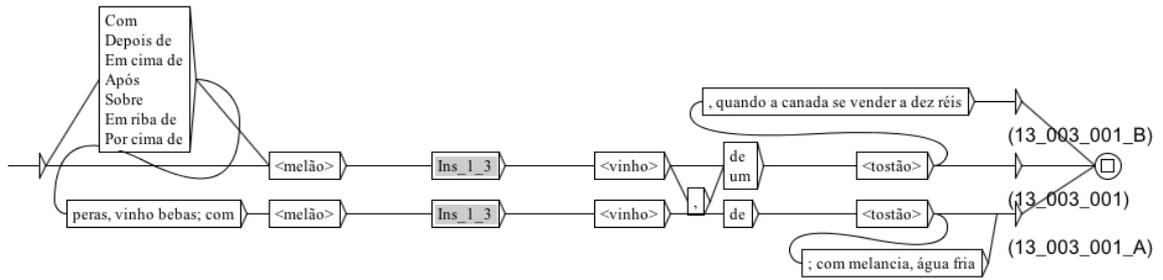


Figura 171: Grafo: Com melão vinho de tostão

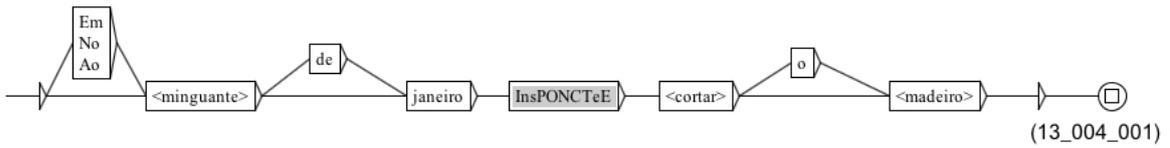


Figura 172: Grafo: No minguante de janeiro, corta o madeiro

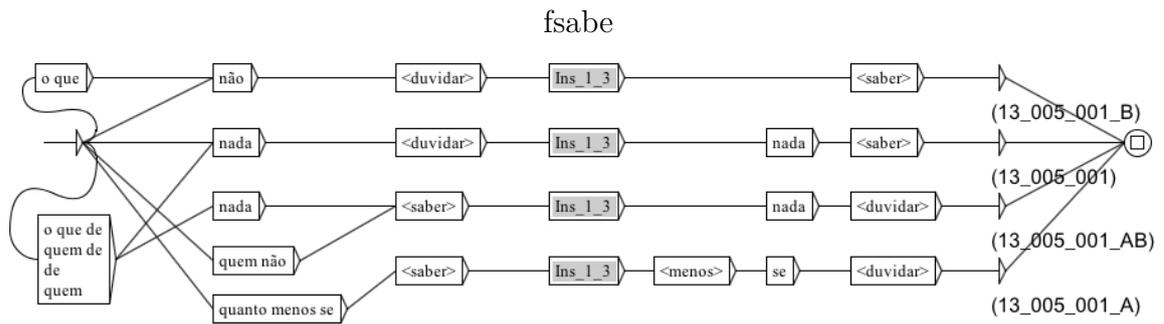


Figura 173: Grafo: Quem de nada duvida nada sabe

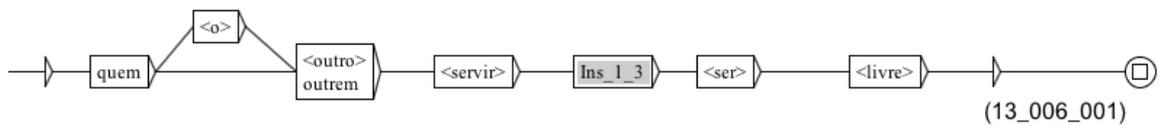


Figura 174: Grafo: Quem outrem serve não é livre

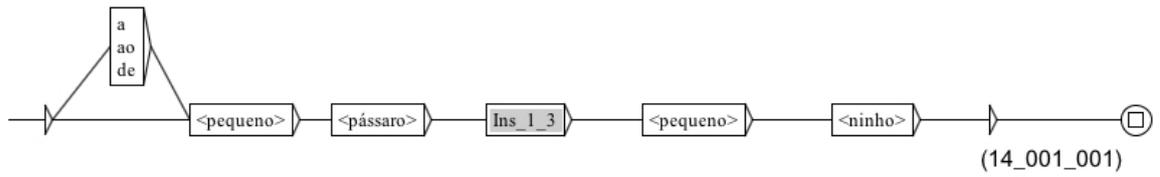


Figura 175: Grafo: A pequeno passarinho pequeno ninho

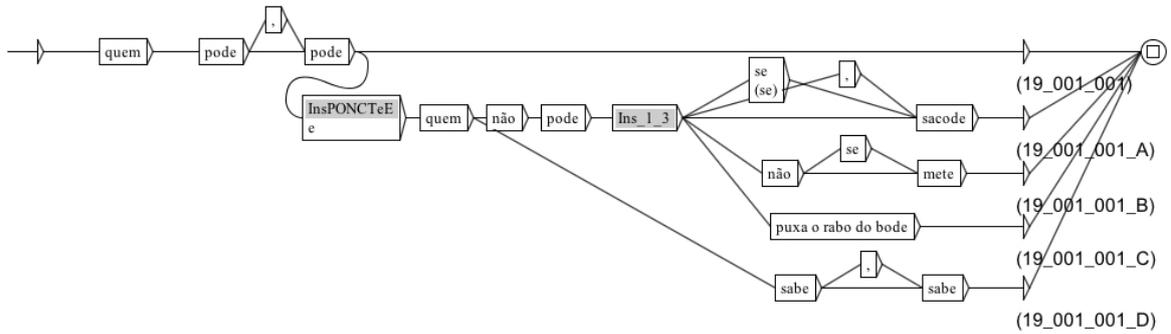


Figura 176: Grafo: Quem pode pode

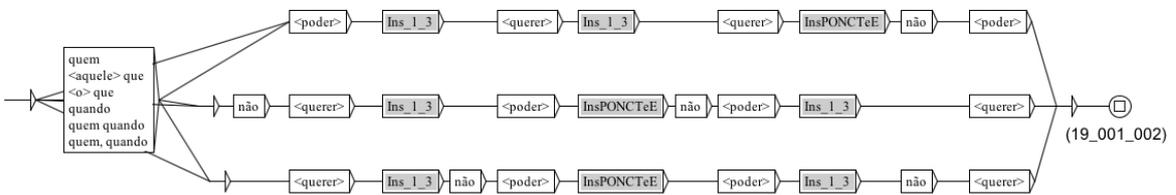


Figura 177: Grafo: Quem pode e não quer quando quer não pode

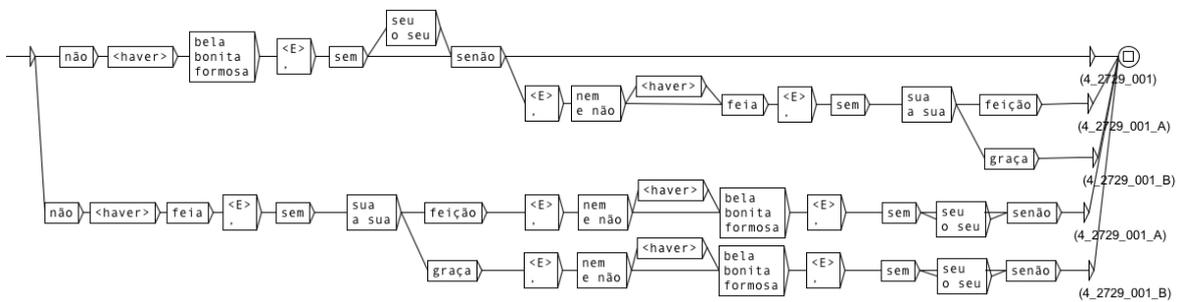


Figura 178: Grafo: Não há bela sem senão

.5 Questionários

Provérbios usuais

Muito obrigada pela sua participação.

Este questionário foi construído no âmbito de um projeto de investigação em Ciências da Linguagem e tem por objetivo determinar a usabilidade de alguns provérbios portugueses. O questionário deve demorar cerca de 15 minutos a realizar. Note que não há respostas certas ou erradas, interessa-nos apenas que nos indique se conhece e/ou usa os provérbios (ou uma das suas variantes) que lhe apresentamos. Este inquérito é absolutamente anónimo e as informações recolhidas destinam-se apenas a processamento estatístico.

***Obrigatório**

1. Sexo (m/f) *

Marcar apenas uma oval.

- Masculino
- Feminino
- Prefiro não dizer

2. Idade *

Marcar apenas uma oval.

- menos de 18
- entre 18 a 30 anos
- entre 30 a 50 anos
- mais de 50 anos

3. Escolaridade *

Marcar apenas uma oval.

- Ensino Básico
- Ensino Secundário
- Ensino Superior
- Outra: _____

4. Nacionalidade *

5. Distrito de residência *

Marcar apenas uma oval.

- Aveiro
- Beja
- Braga
- Bragança
- Castelo Branco
- Coimbra
- Évora
- Faro
- Guarda
- Leiria
- Lisboa
- Portalegre
- Porto
- Santarém
- Setúbal
- Viana do Castelo
- Vila Real
- Viseu
- Açores - Angra
- Açores - Horta
- Açores - Ponta Delgada
- Madeira - Funchal

6. Área de residência *

Marcar apenas uma oval.

- Rural
- Urbana

Questionário

Marque os provérbios seguintes indicando NÃO CONHEÇO; CONHEÇO MAS NÃO USO; CONHEÇO E USO

7. [ID-001] A descer todos os santos ajudam. *

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

8. **[ID-065] Só conhecemos a Deus quando nos foge e ao diabo, quando vem ter conosco. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

9. **[ID-021] O seguro morreu de velho. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

10. **[ID-087] O que se aprende no berço dura até à sepultura. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

11. **[ID-049] O tempo é dinheiro. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

12. **[ID-086] Cedo vos veja como desejo. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

13. **[ID-027] Quem tem unhas é que toca guitarra. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

14. **[ID-044] O meu é meu, o teu é nosso. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

15. **[ID-012] São mais as vozes que as nozes. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

16. **[ID-046] Um homem prevenido vale por dois. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

17. **[ID-096] Cabra manca não tem sesta. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

18. **[ID-055] Ao que erra, perdoa-lhe uma vez, não duas ou três. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

19. **[ID-013] Quem sabe sabe. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

20. **[ID-090] Almocreves samos, na mesma estrada andamos. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

21. **[ID-014] A preguiça morreu de sede à beira da água. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

22. **[ID-032] Um santo não pode estar em dois altares. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

23. **[ID-076] O dever antes do prazer. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

24. **[ID-099] O que outrem sua, muito amansa. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

25. **[ID-080] Estamos no tempo dos dois irmãos: um limpa o nariz, outro bufa às mãos. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

26. **[ID-002] Não se apanham trutas a bragas enxutas. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

27. **[ID-022] A esperança é sempre a última a morrer. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

28. **[ID-031] Não sujes a água que hás-de beber. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

29. **[ID-034] Nem só de pão vive o homem. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

30. **[ID-043] Morra o gato, morra farto. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

31. **[ID-094] A mulher que perde a vergonha nunca a cobra. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

32. [ID-009] Quem não tem dinheiro não tem vícios. *

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

33. [ID-058] Do homem assinalado, sê desconfiado: Deus que o assinalou algum defeito lhe achou. *

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

34. [ID-048] Deitar cedo e cedo erguer dá saúde e faz crescer. *

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

35. [ID-068] Antes pobre e honrado do que rico e ladrão. *

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

36. [ID-091] Quem para diante não olha, atrás fica. *

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

37. [ID-019] O tempo voa. *

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

38. **[ID-064] Da feia e da formosa, a mais proveitosa. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

39. **[ID-077] Debaixo duma roupa feia, há um belo bebedor. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

40. **[ID-029] Para vilão, vilão e meio. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

41. **[ID-006] Diz-me com quem andas, dir-te-ei quem és. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

42. **[ID-067] A mulher andeja diz de todos e todos dela. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

43. **[ID-060] Amigo paga a amigo. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

44. **[ID-070] Como canta o galo velho, assim cantará o novo. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

45. **[ID-082] Moinha de Abril vale por mil. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

46. **[ID-051] O que vai ao tear, ao forno e à frigideira cheira. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

47. **[ID-093] Há lágrimas que nos enganam a nós mesmos depois de terem enganado os outros. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

48. **[ID-025] Há gente para tudo. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

49. **[ID-030] No tempo quente, refresca o ventre. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

50. **[ID-020] A morte não escolhe idades. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

51. **[ID-033] Quem mais fala mais erra. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

52. **[ID-059] Noite perdida nunca é restituída. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

53. **[ID-035] Mulher honrada não tem ouvidos. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

54. **[ID-028] Pensa duas vezes antes de falar. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

55. **[ID-008] Todos os caminhos vão dar a Roma. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

56. **[ID-018] Antes pouco que nada. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

57. **[ID-069] Azul e verde, ranho na parede. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

58. **[ID-097] Por temor, não percas honor. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

59. **[ID-011] Quem vê caras não vê corações. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

60. **[ID-039] O tempo tudo cura. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

61. **[ID-062] Quando não há trigo, come-se de rala. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

62. **[ID-085] Arrenego do homem de muitos barretes. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

63. **[ID-081] Querer desculpar a asneira é asneiar doutra maneira. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

64. **[ID-100] Se queres ver o vilão, mete-lhe a vara na mão. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

65. **[ID-083] Pelo Natal, sol; pela Páscoa, carvão. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

66. **[ID-005] Um mal nunca vem só. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

67. **[ID-042] Quem avisa, amigo é. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

68. **[ID-010] A galinha da vizinha é sempre melhor que a minha. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

69. **[ID-071] Espantalho sem avesso, nem direito. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

70. **[ID-079] Bom gado e bom lavor são minas do agricultor. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

71. **[ID-063] Casa de pobre, tacho de cobre. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

72. **[ID-066] Ainda que vistais o Maio de seda, Maio se queda. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

73. **[ID-057] Dormir com janela aberta, constipação quase certa. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

74. **[ID-072] Livra-te de amigo lisonjeiro e de frade sem mosteiro. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

75. **[ID-089] Que chova, que não chova, meu amo que coma. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

76. **[ID-004] Não há duas sem três. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

77. **[ID-056] Galinha e mulher, não se deixam passear. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

78. **[ID-015] Mais vale tarde do que nunca. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

79. **[ID-074] Não tenhas cão se és tu que ladras. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

80. **[ID-088] Dum homem pequeno não esperes grande coisa. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

81. **[ID-073] Ovo brando comer embaraçado. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

82. **[ID-052] Gota é mal de rico; cura-se fechando o bico. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

83. **[ID-098] Pois que o bispo anda de catana, use de mitra o soldado. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

84. **[ID-024] Hoje por mim, amanhã por ti. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

85. **[ID-040] Na hora do aperto é que se conhecem os amigos. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

86. **[ID-038] Um crime não justifica outro. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

87. **[ID-045] Antes dobrar que quebrar. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

88. **[ID-047] Esta vida são dois dias. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

89. **[ID-016] Quem vai para o mar avia-se primeiro em terra. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

90. **[ID-036] Antes a lã se perca que a ovelha. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

91. **[ID-084] Para um traidor dois aleivosos. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

92. **[ID-053] Uma túnica emprestada nunca aquece e, se aquece, por pouco tempo é. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

93. **[ID-007] Os homens não se medem aos palmos. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

94. **[ID-075] De calar ninguém se arrepende. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

95. **[ID-017] Há sempre uma primeira vez para tudo. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

96. **[ID-078] A piedade sustenta e conserva o que a cruzeza acaba e destrui. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

97. **[ID-095] O que não sabe gozar da ventura, quando ela bate à porta, não deve queixar-se de ser infeliz. ***

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

98. **[ID-061] Quando fizeres o bem, oculta-o; quando te fizerem o bem, divulga-o. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

99. **[ID-041] Não se deve cuspir no prato onde se comeu. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

100. **[ID-092] Viúva rica com um chora com outro repenica. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

101. **[ID-037] Não devas a quem deveu, nem sirvas a quem serviu. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

102. **[ID-050] Não degenera quem sai aos seus. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

103. **[ID-054] Se queres o velho menino, em cima de doce dá-lhe vinho. ***

Marcar apenas uma oval.

- Não conheço
 Conheço mas não uso
 Conheço e uso

104. [ID-026] O sol quando nasce é para todos. *

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

105. [ID-003] Depois da tempestade vem a bonança. *

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

106. [IID-023] Santos de casa não fazem milagres. *

Marcar apenas uma oval.

- Não conheço
- Conheço mas não uso
- Conheço e uso

Muito obrigada pela sua participação!

Com tecnologia



Formulário para estimativa de tempo de resposta de jogos proverbiais

Muito obrigada pela sua participação.

Este questionário de teste foi construído no âmbito de um projeto de investigação em Ciências da Linguagem e tem por objetivo estimar o tempo de resposta de 10 jogos de provérbios, de forma a que possamos determinar a quantidade de jogos a inserir no formulário final. Note que não há respostas certas ou erradas, interessa-nos apenas que seleccione a/as palavras em falta em cada provérbio que lhe apresentamos e que indique a hora precisa a que iniciou e finalizou o conjunto de jogos. Este inquérito é absolutamente anónimo e as informações recolhidas destinam-se apenas a processamento estatístico.

***Obrigatório**

1. **Insira a hora, por favor. ***

Exemplo: 08:30

2. **[QT 1] A esperança é sempre a última a _____ ***

Marcar apenas uma oval.

- morrer
- crescer
- desaparecer
- acabar

3. **[QT 2] Quem tem unhas _____ ***

Marcar apenas uma oval.

- é que toca bandola
- é que toca guitarra
- não as roa
- é que toca viola

4. **[QT 3] O _____ quando nasce é para todos. ***

Marcar apenas uma oval.

- arco-íris
- dia
- rio
- sol

5. [QT 4] A galinha da vizinha _____ *

Marcar apenas uma oval.

- é sempre pior que a minha
- é mais magra que a minha
- é sempre melhor que a minha
- põe mais ovos do que a minha

6. [QT 5] São mais as vozes que as _____ *

Marcar apenas uma oval.

- nozes
- cantigas
- doses
- marimbas

7. [QT 6] Pensa duas vezes _____ *

Marcar apenas uma oval.

- antes de calar
- antes de abalar
- antes de jurar
- antes de falar

8. [QT 7] Morra Marta _____ *

Marcar apenas uma oval.

- mas morra regalada
- morra quarta
- morra farta
- pois viveu farta

9. [QT 8] Um homem _____ vale por dois. *

Marcar apenas uma oval.

- prevenido
- prudente
- desprevenido
- atrevido

10. [QT 9] Antes _____ que quebrar. *

Marcar apenas uma oval.

- dobrar
- vergar
- sujeitar
- aguentar

11. [QT 10] Deitar cedo e cedo erguer _____ *

Marcar apenas uma oval.

- alonga o viver
- dá saúde para crescer
- cria carne e sebo
- dá saúde e faz crescer

12. Insira a hora, por favor. *

Exemplo: 08:30

13. Quer fazer algum comentário ou sugestão adicional? A sua opinião é muito importante para nós.

14. Se desejar receber mais informação sobre este projeto, deixe o seu contacto de e-mail

15. Se desejar receber mais informação sobre este projeto, deixe o seu contacto de e-mail

Com tecnologia

 Google Forms

Questionário de avaliação da qualidade de jogos proverbiais

Este questionário foi construído no âmbito de um projeto de investigação em Ciências da Linguagem e tem por objetivo avaliar a qualidade de um conjunto de jogos com provérbios com vista a vir a inseri-los numa plataforma virtual de diagnóstico e terapia de algumas patologias da linguagem. O questionário deve demorar cerca de 15 minutos a realizar. Este inquérito é absolutamente anónimo e as informações recolhidas destinam-se apenas a processamento estatístico.

A sua participação é muito importante para nós.

Obrigada por participar neste questionário.

***Obrigatório**

1. Distrito de residência habitual *

Marcar apenas uma oval.

- Aveiro
- Beja
- Braga
- Bragança
- Castelo Branco
- Coimbra
- Évora
- Faro
- Guarda
- Leiria
- Lisboa
- Portalegre
- Porto
- R. A. Açores
- R. A. Madeira
- Santarém
- Setúbal
- Viana do Castelo
- Vila Real
- Viseu
- Resido fora de Portugal

2. Escolaridade *

Marcar apenas uma oval.

- Ensino Básico
- Ensino Secundário
- Ensino Superior
- Outra: _____

3. Idade *

Marcar apenas uma oval.

- menos de 18 anos
- entre 18 a 30 anos
- entre 31 a 50 anos
- mais de 50 anos

4. Nacionalidade *

Marcar apenas uma oval.

- Portuguesa
- Outra: _____

5. Sexo (m/f) *

Marcar apenas uma oval.

- Masculino
- Feminino
- Prefiro não dizer

Jogos proverbiais

Selecione uma das opções apresentadas de forma a completar o provérbio.

6. [Q-025] O _____ não faz o monge. *

Marcar apenas uma oval.

- vestuário
- capote
- hábito
- costume

7. [Q-055] Quem sabe _____. *

Marcar apenas uma oval.

- não estranha
- que se gabe
- ensina
- sabe

8. [Q-095] **Longe da vista, _____.** *

Marcar apenas uma oval.

- longe da razão
- perto da emoção
- longe do coração
- não há quem resista

9. [Q-081] **Laranja, de manhã é ouro, à tarde prata _____.** *

Marcar apenas uma oval.

- e de noite enfastia
- e à noite é lata
- e à noite mata
- e à noite é choro

10. [Q-048] **O silêncio é a alma do _____.** *

Marcar apenas uma oval.

- sacerdócio
- contrato
- sossego
- negócio

11. [Q-041] **Nem tudo o que vem à _____ é peixe.** *

Marcar apenas uma oval.

- armadilha
- rede
- linha
- baila

12. [Q-064] _____, **dá saúde e faz crescer.** *

Marcar apenas uma oval.

- Deitar tarde e tarde erguer,
- Dormir bem e bem comer,
- Deitar cedo e cedo erguer,
- Levantar antes do amanhecer,

13. [Q-098] Não há duas _____.*

Marcar apenas uma oval.

- para quem não refez
- sem três
- sem uma vez
- sem quatro

14. [Q-043] Onde há _____, há fogo.*

Marcar apenas uma oval.

- nuvem
- fumo
- bombeiros
- fumaceira

15. [Q-097] Cada um puxa a brasa _____.*

Marcar apenas uma oval.

- ao seu peixe
- à sua pinha
- à sua sardinha
- para o lado que que convem

16. [Q-100] Em casa de ferreiro, _____.*

Marcar apenas uma oval.

- o espeto é de ferro
- espeto de pau
- há sempre um bom faqueiro
- espeto de madeira

17. [Q-045] A _____ não perdoa.*

Marcar apenas uma oval.

- verdade
- vaidade
- idade
- velhice

18. [Q-044] _____, **leva-as o vento.** *

Marcar apenas uma oval.

- Cantigas
- Palavras
- Formigas
- Canções

19. [Q-023] **Nem tanto ao _____, nem tanto à terra.** *

Marcar apenas uma oval.

- oceano
- mar
- ar
- lago

20. [Q-070] **Azar ao jogo, _____.** *

Marcar apenas uma oval.

- sorte no casamento
- sorte no amor
- azar em tudo
- é para os principiantes

21. [Q-080] **Olho por olho, _____.** *

Marcar apenas uma oval.

- mão por mão
- dente por dente
- ficas zarolho
- gente por gente

22. [Q-035] **As paredes têm _____.** *

Marcar apenas uma oval.

- badana
- orelha
- ouvidos
- tijolo

23. [Q-042] Quem brinca com o _____ queima-se. *

Marcar apenas uma oval.

- carvão
- óleo
- fogo
- jogo

24. [Q-086] Quando mija um português, _____. *

Marcar apenas uma oval.

- mijam dois ou três
- mija claro e fala direito
- não mija só uma vez
- não mija um francês

25. [Q-038] A cavalo dado, não se olha ao _____. *

Marcar apenas uma oval.

- arado
- pedigree
- pigmento
- dente

26. [Q-090] Deus dá o frio _____. *

Marcar apenas uma oval.

- e o Diabo dá o brio
- conforme a roupa
- conforme o arrepio
- conforme a vestimenta

27. [Q-094] Quem não é por mim, _____. *

Marcar apenas uma oval.

- é porque só sabe de si
- é contra mim
- não é por ninguém
- é contra ti

28. [Q-016] _____ não dorme. *

Marcar apenas uma oval.

- Deus
- Zeus
- Mateus
- semideus

29. [Q-018] A _____ acção fica com quem a pratica. *

Marcar apenas uma oval.

- pior
- maldosa
- boa
- má

30. [Q-063] _____ são dois dias. *

Marcar apenas uma oval.

- As festas
- A existência
- A vida
- Esta lida

31. [Q-077] Tal pai, _____. *

Marcar apenas uma oval.

- tal descendente
- tal fruto
- tal filho
- tal neto

32. [Q-003] Os _____ não se medem aos palmos. *

Marcar apenas uma oval.

- líderes
- jovens
- cavalheiros
- homens

33. [Q-046] A _____ tem limites. *

Marcar apenas uma oval.

- paciência
- ciência
- persistência
- pachorra

34. [Q-068] Zangam-se as comadres, _____.

Marcar apenas uma oval.

- acabam-se as crueldades
- vão-se as maldades
- descobrem-se as verdades
- encobrem-se as verdades

35. [Q-009] Santos de casa não fazem _____.

Marcar apenas uma oval.

- padroeiros
- maravilhas
- milagres
- prodígios

36. [Q-019] Três foi a conta que _____ fez.

Marcar apenas uma oval.

- Mateus
- Deus
- Zeus
- português

37. [Q-059] Hoje por mim, _____.

Marcar apenas uma oval.

- hoje sem fim
- amanhã por mim
- amanhã por ti
- amanhã por vós

38. [Q-072] Toma lá, _____.*

Marcar apenas uma oval.

- dá acolá
- dá lá
- dá cá
- leva lá

39. [Q-030] Quem semeia ventos colhe _____.*

Marcar apenas uma oval.

- temporais
- vendavais
- tempestades
- tormentos

40. [Q-047] Cão que ladra não _____.*

Marcar apenas uma oval.

- foge
- morde
- roe
- ferra

41. [Q-079] Para quem é, _____.*

Marcar apenas uma oval.

- bacalhau basta
- carapau basta
- nada basta
- qualquer coisa chega

42. [Q-012] Com o _____, tudo se cura.*

Marcar apenas uma oval.

- amanhã
- tempo
- futuro
- momento

43. [Q-008] O _____ voa. *

Marcar apenas uma oval.

- dinheiro
- vento
- tempo
- cheiro

44. [Q-005] A _____ da vizinha é sempre melhor que a minha. *

Marcar apenas uma oval.

- andorinha
- burrinha
- galinha
- vaca

45. [Q-028] A roupa _____ lava-se em casa. *

Marcar apenas uma oval.

- encardida
- imaculada
- suja
- manchada

46. [Q-069] Quem não chora _____.

Marcar apenas uma oval.

- não come
- mija
- não ama
- não mama

47. [Q-010] Melhor é dobrar que _____.

Marcar apenas uma oval.

- quebrar
- morrer
- abalar
- render

48. [Q-082] **Entre marido e mulher, _____.** *

Marcar apenas uma oval.

- nunca metas a colher
- nunca metas o talher
- nunca se mete o bedelho
- mete sempre a colher

49. [Q-096] _____ **enche a galinha o papo.** *

Marcar apenas uma oval.

- Bago a bago
- Grão a grão
- Comendo pão
- Com o trapo

50. [Q-071] **Quem não deve _____.** *

Marcar apenas uma oval.

- não teve
- não teme
- não receia
- não paga

51. [Q-076] **Quem tudo quer, _____.** *

Marcar apenas uma oval.

- mais consegue
- tudo ganha
- tudo perde
- mais perde

52. [Q-040] **A ocasião faz o _____.** *

Marcar apenas uma oval.

- ladrão
- larpão
- larápio
- coração

53. [Q-073] Quem com ferro fere _____.*

Marcar apenas uma oval.

- não é bom partido
- com ferro será ferido
- com ferro não será ferido
- com ferro será batido

54. [Q-014] O que arde _____.*

Marcar apenas uma oval.

- dura
- cura
- fura
- ajuda

55. [Q-020] O pior _____ é o que não quer ver.*

Marcar apenas uma oval.

- ignorante
- cego
- indouto
- desconhecedor

56. [Q-056] Há sempre uma primeira vez _____.*

Marcar apenas uma oval.

- para tudo
- para perder
- para quem fez
- para o português

57. [Q-084] Quem casa _____.*

Marcar apenas uma oval.

- quer lar
- quer asa
- quer casa
- quer teto

58. [Q-065] O tempo _____.*

Marcar apenas uma oval.

- é fortuna
- é riqueza
- é dinheiro
- é tesouro

59. [Q-029] Parar é _____.*

Marcar apenas uma oval.

- perecer
- fraquejar
- perder
- morrer

60. [Q-087] As moscas apanham-se _____.*

Marcar apenas uma oval.

- com vinagre
- com açúcar
- com mel e não com fel
- com fel e não com mel

61. [Q-007] Mais vale tarde do que _____.*

Marcar apenas uma oval.

- jamais
- cedo
- sempre
- nunca

62. [Q-006] São mais as _____ que as nozes.*

Marcar apenas uma oval.

- doses
- nozes
- cantigas
- marimbas

63. [Q-032] _____ escaldado, de água fria tem medo. *

Marcar apenas uma oval.

- Gato
- Pato
- Peixe
- Felino

64. [Q-015] Quem está mal que se _____. *

Marcar apenas uma oval.

- grude
- refugie
- mude
- cale

65. [Q-021] A vingança serve-se _____. *

Marcar apenas uma oval.

- gelada
- fria
- quente
- mansa

66. [Q-031] Quanto mais se tem, mais se _____. *

Marcar apenas uma oval.

- ambiciona
- economiza
- precisa
- quer

67. [Q-001] A descer todos os _____ ajudam. *

Marcar apenas uma oval.

- santos
- anjos
- diabos
- demónios

68. [Q-066] Quem sai aos seus _____.*

Marcar apenas uma oval.

- não desespera
- não degenera
- não desonera
- degenera

69. [Q-083] O fruto proibido _____.*

Marcar apenas uma oval.

- nunca é o mais apetecido
- é o mais apetecido
- é o mais querido
- é o mais cobiçado

70. [Q-033] Burro velho não aprende _____.*

Marcar apenas uma oval.

- trava-línguas
- latim
- idioma
- línguas

71. [Q-057] O seguro _____.*

Marcar apenas uma oval.

- morreu adulto
- morreu de antigo
- morreu centenário
- morreu de velho

72. [Q-092] Quanto mais alto se sobe, _____.*

Marcar apenas uma oval.

- menor é a queda
- maior é o trambolhão
- mais depressa se cai
- maior é a queda

73. [Q-099] Quem espera, _____.*

Marcar apenas uma oval.

- sempre alcança
- desespera
- tem de ter paciência
- colhe

74. [Q-089] Os olhos comem mais _____.*

Marcar apenas uma oval.

- que a barriga
- que o ventre
- que a pança
- do que os pardais

75. [Q-036] Palavra puxa _____.*

Marcar apenas uma oval.

- argumento
- frase
- palavra
- carroça

76. [Q-091] O prometido _____.*

Marcar apenas uma oval.

- está em dívida
- é devido
- é cumprido
- é garantido

77. [Q-093] A palavras loucas, _____.*

Marcar apenas uma oval.

- orelhas moucas
- ouvidos surdos
- orelhas taralhucas
- ensinanças poucas

78. [Q-067] Não faças aos outros _____.*

Marcar apenas uma oval.

- o que não queres que te façam a ti
- nada ruim
- o que não gostas
- o que não te fazem a ti

79. [Q-050] A _____ não tem preço.*

Marcar apenas uma oval.

- qualidade
- saúde
- lucidez
- cultura

80. [Q-011] Quem tem _____ é que toca guitarra.*

Marcar apenas uma oval.

- garra
- unhas
- palheta
- cunhas

81. [Q-049] A _____ é o melhor tempero.*

Marcar apenas uma oval.

- mostarda
- fome
- gana
- salsa

82. [Q-054] Quem vê caras _____.*

Marcar apenas uma oval.

- não vê taras
- não vê corações
- não vê brasões
- vê ilusões

83. [Q-075] Enquanto há vida, _____.*

Marcar apenas uma oval.

- há aliança
- há cobrança
- há lembrança
- há esperança

84. [Q-013] A curiosidade matou o _____.*

Marcar apenas uma oval.

- frade
- sapo
- rato
- gato

85. [Q-024] Não há rosa sem _____.*

Marcar apenas uma oval.

- picos
- estames
- espinhos
- botão

86. [Q-039] Apanha-se mais depressa um _____ que um coxo.*

Marcar apenas uma oval.

- franco
- falso
- mentiroso
- idoso

87. [Q-022] Abril, _____ mil.*

Marcar apenas uma oval.

- chuvas
- águas
- mágoas
- marés

88. [Q-027] Quem anda à chuva _____.*

Marcar apenas uma oval.

- molha-se
- encharca-se
- banha-se
- lava-se

89. [Q-053] Todos os caminhos _____.*

Marcar apenas uma oval.

- vão dar ao rio
- vão dar a Oklahoma
- levam a Roma
- vão dar a Roma

90. [Q-085] As más notícias _____.*

Marcar apenas uma oval.

- vêm sem pressa
- vêm devagar
- chegam depressa
- chegam sem demora

91. [Q-004] Quem não tem _____ não tem vícios.*

Marcar apenas uma oval.

- benefícios
- fortuna
- dinheiro
- ofícios

92. [Q-002] Depois da tempestade vem a _____.*

Marcar apenas uma oval.

- bonança
- felicidade
- tranquilidade
- esperança

93. [Q-058] A esperança _____.*

Marcar apenas uma oval.

- é verde
- é sempre a última a morrer
- é a última a nascer
- é sempre a primeira a morrer

94. [Q-051] Um mal _____.*

Marcar apenas uma oval.

- nunca vem desamparado
- nunca vem acompanhado
- nunca vem só
- vem sempre sozinho

95. [Q-078] Quem cala _____.*

Marcar apenas uma oval.

- não desmente
- consente
- não mente
- permite

96. [Q-017] À primeira cavadela, _____.*

Marcar apenas uma oval.

- feijoca
- mandioca
- agúdias
- minhoca

97. [Q-052] Diz-me com quem andas, _____.*

Marcar apenas uma oval.

- dir-te-ei quem és
- dir-te-ei as manhas que tens
- dir-te-ei quem serás
- e assim serão as demandas

98. [Q-037] A _____ é a mãe da ciência. *

Marcar apenas uma oval.

- sapiência
- vivência
- experiência
- razão

99. [Q-074] Quem está vivo _____.

Marcar apenas uma oval.

- sempre aparece
- sempre se mostra
- é porque não morreu
- sempre comparece

100. [Q-034] _____ não puxa carroça. *

Marcar apenas uma oval.

- Feixe
- Peixe
- Burro
- Cardume

101. [Q-060] O sol quando nasce _____.

Marcar apenas uma oval.

- é para quem merece
- não é para todos
- é para todos
- vem de Este

102. [Q-088] Com uma mentira _____.

Marcar apenas uma oval.

- se apanha a realidade
- se apanha a falsidade
- se apanha uma maldade
- se apanha uma verdade

103. [Q-026] A _____ é boa conselheira. *

Marcar apenas uma oval.

- amiga
- vigília
- noite
- almofada

104. [Q-062] Um homem prevenido _____.*

Marcar apenas uma oval.

- vale por dois
- jamais será vencido
- vale por muitos
- deve ser contido

105. [Q-061] Quem te avisa, _____.*

Marcar apenas uma oval.

- teu amigo é
- teu amigo não é
- não banaliza
- teu aliado é

Questionário de avaliação da qualidade de jogos proverbiais

Nesta breve secção pretendemos colocar-lhe 5 questões sobre o jogo apresentado anteriormente.

Em poucos segundos terá terminado este questionário.

106. 2 - A solução do exercício estava entre as opções indicadas? *

Marcar apenas uma oval.

- Sempre
- A maioria das vezes
- Poucas vezes
- Nunca

107. 1 - Conhecia todos os provérbios que lhe foram apresentados? *

Marcar apenas uma oval.

- Sempre
- A maioria das vezes
- Poucas vezes
- Nunca

108. **4 - Encontrou casos em que havia mais do que uma resposta correta? ***

Marcar apenas uma oval.

- Sempre
- A maioria das vezes
- Poucas vezes
- Nunca

109. **3 - Mesmo quando não conhecia o provérbio utilizado, conseguiu completá-lo facilmente com base nas opções de resposta apresentadas? ***

Marcar apenas uma oval.

- Sempre
- A maioria das vezes
- Poucas vezes
- Nunca
- Não se aplica: conhecia todos os provérbios

110. **Se desejar receber mais informação sobre este projeto, deixe-nos o seu contacto de e-mail**

111. **5 - Usa os provérbios apresentados? ***

Marcar apenas uma oval.

- Sim, uso todos
- Sim, uso a maioria deles
- Sim, mas apenas uso alguns
- Não, não uso nenhuns

112. **Quer fazer algum comentário ou sugestão adicional? A sua opinião é muito importante para nós.**

Com tecnologia

 Google Forms

As intenções comunicativas no uso de provérbios.

Muito obrigada pela sua participação, que é muito importante para nós. Este questionário foi construído no âmbito de um projeto de investigação em Ciências da Linguagem e tem por objetivo aferir/determinar com que intenções comunicativas são usados os provérbios. Note que não há respostas certas ou erradas, interessa-nos apenas que leia atentamente cada um dos excertos apresentados (retirados da internet tal e qual estão) e que identifique no texto os elementos ou expressões sobre os quais o provérbio é dito. Deverá, então, selecionar esses elementos ou expressões, copiá-los e colá-los para o campo destinado para o efeito. O questionário deve demorar cerca de 30 minutos a realizar. Os dados fornecidos são absolutamente anónimos e confidenciais e as informações recolhidas destinam-se apenas a processamento estatístico.

***Obrigatório**

1. Sexo (m/f) *

Marcar apenas uma oval.

- masculino
- feminino
- prefiro não dizer

2. Idade *

Marcar apenas uma oval.

- menos de 18 anos
- entre 18 e 30 anos
- entre 31 e 50 anos
- mais de 50 anos

3. Escolaridade *

Marcar apenas uma oval.

- Ensino Básico
- Ensino Secundário
- Ensino Superior
- Outra: _____

4. Nacionalidade *

5. Distrito de residência *

Marcar apenas uma oval.

- Aveiro
- Beja
- Braga
- Bragança
- Castelo Branco
- Coimbra
- Évora
- Faro
- Guarda
- Leiria
- Lisboa
- Portalegre
- Porto
- Santarém
- Setúbal
- Viana do Castelo
- Vila Real
- Viseu
- R.A. Madeira
- R.A. Açores
- Vivo fora de Portugal

6. Área de residência *

Marcar apenas uma oval.

- Rural
- Urbana

Exemplo

Nesta secção, apresentamos-lhe um exemplo do tipo de exercício que irá realizar no questionário. O provérbio alvo deste excerto é #_Cão que ladra não morde_#. Os provérbios encontram-se assinalados da seguinte forma #_PROVÉRBIO_#.

A primeira parte do exercício consiste na leitura de cada um dos excertos que lhe são apresentados, como, por exemplo, o que se encontra abaixo.

(...) "um dia o Pedro descobriu que o sr. Joaquim quando saia da sua loja, vinha carregado com muitos sacos cheios de coisas. E que os levava para as crianças do orfanato e que todas gostavam muito dele e ele era muito atencioso para todas. A conclusão que tiveram foi que o sr. Joaquim pode ralhar muito com nosco mas tambem

é para o nosso próprio bem e que afinal tem um coração muito bom.

Ou seja "#_CÃO QUE LADRA NÃO MORDE_#".

Esta história é para crianças um pouco mais velhas, pois é uma história um pouco maior, mas ensina a elas, que não devemos julgar as pessoas por aquilo que elas aparentam, porque por fora elas podem parecer assim, mas depois lá no fundo são muito boas pessoas."

in <https://carolineptai.webnode.pt/news/historia-%22-c%C3%A3o-que-ladra-n%C3%A3o-morde%21%22/>

De seguida, pedimos-lhe que identifique o(s) elemento(s) do texto sobre o/os qual/quais o provérbio é dito. Copie-os e cole-os em "Coloque aqui a sua resposta".

Coloque aqui a sua resposta: (Neste caso, o provérbio é dito a propósito de:)

"A conclusão que tiveram foi que o sr. Joaquim pode ralhar muito com nosco mas também é para o nosso próprio bem e que afinal tem um coração muito bom."

OU

"não devemos julgar as pessoas por aquilo que elas aparentam, porque por fora elas podem parecer assim, mas depois lá no fundo são muito boas pessoas."

Depois, deverá selecionar uma das seguintes opções que lhe são apresentadas.

(Observe abaixo o exemplo ilustrativo.)

7. O que o provérbio diz sobre esses elementos é:

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

Neste caso, selecionaríamos a opção "positivo", uma vez que o provérbio é usado para realçar que o sr. Joaquim, apesar de ralhar com eles, não é capaz de lhes fazer mal, tendo ainda se mostrado um homem de bom coração. Se escolher as opções "não sei o que responder" ou "não é possível [...] determinar...", digite

QUALQUER COISA no campo de texto, apenas para o validar (essa parte da resposta será ignorada).

Na próxima secção, apresentamos-lhe o questionário. Ser-lhe-ão mostrados 10 provérbios diferentes. Cada provérbio, por sua vez, estará associado a 3 excertos diferentes. Pedimos-lhe que siga as instruções.
Muito obrigada!

Passe para a pergunta 8.

Questionário

Leia atentamente cada um dos excertos apresentados e identifique o elemento ou elementos sobre o/os qual/quais o provérbio é dito. Deverá copiar esse/s elemento/s e colá-lo/s no local destinado para o efeito. Pedimos-lhe ainda que, seguidamente, selecione a opção que considerar mais adequada sobre a intenção comunicativa desse provérbio.

P1 - ID/017-1: A vida são dois dias.

Por favor leia o excerto.

(...) "Classe Optimist Juvenil, sendo esta deslocação muito importante para o desenvolvimento desportivo e social dos atletas. No ginásio AquaGym o programa de festividades carnavalescas inclui duas aulas especiais em pleno Dia de Entrudo, nomeadamente a aula de spinning às 9h30 e a aula de zumba às 10h30, onde os utentes serão convidados a trazer um disfarce sob o tema "América". No dia 17, encerraremos as festividades da época, com uma festa de "Enterro do Osso" aberta a todos os sócios do clube, com entrada livre na sede social. Sob o tema The Great Gatsby Party e com o dress code de anos 20, a festa terá animação a cargo do DJ Kaly. As portas abrem às 21h00 e, aos foliões presentes, estão prometidos prémios para o melhor disfarce bem como algumas surpresas. Razões não faltam para aproveitar esta divertida época no seu clube, já que #_A VIDA SÃO DOIS DIAS_# e o Carnaval são três."

In <http://www.tribunadamadeira.pt/2018/02/09/a-vida-sao-dois-dias-e-o-carnaval-sao-tres/>

8. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

9. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P1 - ID/017-2: A vida são dois dias.

Por favor leia o excerto.

(...) “Este belíssimo carnaval é vivido a 100% pelos torreenses que começam um ano antes a preparar tudo com a ajuda de pessoas que vêm de arredores. Muitos indivíduos vêm de propósito a Torres Vedras nesta altura do ano para poder saber em que é que este Carnaval consiste e o porquê das pessoas afirmam ser o melhor Carnaval de Portugal.

O ano passado o tema foi figuras e figurões, já este ano é Brinquedos e Brincadeiras.

Este carnaval tem uma longa história e uma longa tradição e não é ao acaso que é considerado o carnaval mais português de Portugal. A população da zona de Torres Vedras vive este evento a 100%, é algo que lhes está no sangue e que é passado de geração em geração.

Como #_A VIDA SÃO DOIS DIAS_# e o Carnaval de Torres são seis, preparem-se para vir já amanhã ao primeiro dia do Carnaval de Torres Vedras.”

In <https://torresvedrasweb.pt/a-vida-sao-2-dias-e-o-carnaval-de-torres-sao-6/>

10. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

11. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P1 - ID/017-3: A vida são dois dias.

Por favor leia o excerto.

(...) "Sorri, que #_A VIDA SÃO DOIS DIAS_# e não podes passar, nem um dia, a esconder esse sorriso tão bonito que tens. Não permitas que um dia chuvoso te faça ficar de mal humor. Desfruta de todos os momentos, independentemente da chuva que estiver a cair, com este super guarda-chuvas. Características:- Medidas do guarda-chuvas fechado: 5x18 cm (aprox.)

- Diâmetro do guarda-chuvas aberto: 108 cm
- Material: Poliéster e 7 varas reforçadas.
- Abertura manual.
- Windproof reforçado e anti vento."

in <https://www.fnac.pt/Guarda-Chuva-Sorri-A-Vida-Sao-Dois-Dias-Mr-Wonderful/a960884#>

12. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

13. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P2 - ID/019-1: Abril, águas mil.

Por favor leia o excerto.

“Em Abril águas mil? Parece que sim.

Diz o povo que em #_ABRIL ÁGUAS MIL_#. Nem sempre é assim, nem as previsões para um mês inteiro têm fiabilidade suficiente para darmos agora uma resposta concreta. No entanto, a avaliar pelas previsões que estão disponíveis para esta semana e para a seguinte, parece que se pode vir a cumprir o ditado.

A chuva e até a neve na Serra da Estrela vão continuar a marcar presença nos próximos dias ou semanas. Claro que vão aparecer alguns dias de sol e temperaturas amenas pelo meio, mas a maioria será com chuva e algum frio.

Continuará a nevar na Serra da Estrela quando houver precipitação, serão mais os dias com neve do que os em que seja apenas chuva. Teremos uma Primavera com um bom manto branco que se vai aguentar durante muito tempo.

Como é habitual nesta época do ano, a precipitação será mais através de” (...)

in <https://vitorbaiameteo.pt/2018/04/02/abril-aguas-mil-parece-sim/>

14. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

15. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P2 - ID/019-2: Abril, águas mil.

Por favor leia o excerto.

(...) “Todas as antigas civilizações se semearam, cultivaram, disseminaram junto ou entre grandes rios (Tigre, Eufrates, Ganges, Nilo, Danúbio, Guadalquivir, Douro, Tejo, Guadiana, Sado, Mondego, entre tantos outros).

Mas sem chuva a trazer de volta água à terra que pisamos, a miséria instala-se. Se tarda, se demora, se se ausenta, quebrando os ritmos sazonais anuais, seculares, logo se instalam severas preocupações com a morte a substituir o mar a jusante.

Quase todos ouvimos o ditado popular “#_EM ABRIL, ÁGUAS MIL_#”. Quase nenhum de nós o questiona. É um saber feito de tempo, saber tácito feito de regularidade passada que teima em se repetir anualmente, pouco tempo depois do equinócio (da primavera) em regiões acima da nossa latitude até ao círculo polar ártico.

Aquele ditado não existe só em terras lusas. Que se encontre também em Espanha não é de espantar. Mas que faça parte da “sabedoria popular” por essa Europa acima, isso já levanta algum” (...)

in <http://dererummundi.blogspot.com/2012/04/abril-aguas-mil.html>

16. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

17. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo que negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P2 - ID/019-3: Abril, águas mil.

Por favor leia o excerto.

“Quase todos conhecemos o ditado popular “#_EM ABRIL, ÁGUAS MIL_#”. Já o interiorizamos e ninguém o questiona. Faz parte da sabedoria popular. É um saber feito de tempo que teima em repetir-se anualmente, pouco tempo depois do equinócio (da primavera) em regiões acima da nossa latitude até ao círculo polar ártico.

O fenómeno meteorológico que alimenta o provérbio assenta num aumento de luminosidade solar incidente, de forma progressivamente mais perpendicular a partir do equinócio da primavera (no hemisfério norte). Isto provoca um aumento progressivo da temperatura do solo, o que, por sua vez, causa evaporação da água (devido a outonos e invernos mais chuvosos – como foi o caso este ano) retida e presente nos interstícios da terra.

Por isso abril é sinónimo de chuva. É um daqueles meses que me deixa com sentimentos contraditórios, se por um lado é um mês agradável, os dias começam a ficar mais longos, o” (...)

in <https://narrativadiaria.blogs.sapo.pt/abril-aguas-mil-149293>

18. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

19. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P3 - ID/065-1: Os olhos comem mais que a barriga.

Por favor leia o excerto.

"Um espaço amplo, tranquilo, onde tanto para jantar a 2 como para grupos, família ou amigos, casamentos ou eventos e até mesmo um simples cafezinho ao ar livre, mais do que excelente. Noites de espectáculos de Verão, noites de Fado de inverno, acede a todos os gostos e falando da comida, só nas entradas já #_OS OLHOS COMEM MAIS QUE A BARRIGA_#, não vale a pena mencionar os excelentes pratos, e de tudo a minha favorita "Sobremesa do Foral". Recomendadíssimo
Pergunte a KarynaMendes acerca de Restaurante O Foral"

in https://www.tripadvisor.pt/ShowUserReviews-g2626749-d8146370-r599465290-Restaurant_O_Foral-Algoz_Faro_District_Algarve.html

20. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

21. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P3 - ID/065-2: Os olhos comem mais que a barriga.

Por favor leia o excerto.

"Cowboy Solitario disse:

Sim já foi discutido várias vezes esse tipo de tema, do genero haver estações do ano, noite e dia, etc...

A questão é que alguns defendem que isso não interessa nada e não traz nada de novo para o jogo. Só vai fazer para o jogo ser lento.

Eu também acho que o jogo deveria sofrer alterações a esse aspecto, seria mais interessante pelo menos se passasse para noite.

nao tráz nada de novo é verdade...mas torna-se bastante mais engraçado..=D dava mais aspecto ao jogo..alias sao pequenos pormenores que fazem a diferença..nunca se esqueçam que #_OS OLHOS COMEM MAIS QUE A BARRIGA_# por isso o aspecto é essencial :D:D"

in <https://forum.the-west.com.pt/index.php?threads/sugestao.4456/>

22. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

23. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P3 - ID/065-3: Os olhos comem mais que a barriga.

Por favor leia o excerto.

"Olha que eu era gaja para experimentar quase todos, à excepção dos ratinhos.... Mas pensando melhor acho que até esses iam, também comemos codornizes (que raio de comparação)

[...]

LadyVih 17.03.2016 09:18
MaravilhososImage...

Maria Amélia 16.03.2016 13:26
Coisas demasiado estranhas para mim *-*

M.J. 16.03.2016 18:13
diz que é uma questão de sabor.

Just_Smile 16.03.2016 13:28
Iguarias que dispenso perfeitamente experimentar :P

M.J. 16.03.2016 18:13
pode ser bom...

Just_Smile 16.03.2016 18:20
Nunca ouviste dizer que #_OS OLHOS COMEM MAIS QUE A BARRIGA_#? :P

O Cliente 16.03.2016 14:08
ai ai, esse estômago deu a volta ao meu ahah

M.J. 16.03.2016 18:13
AHAHAHAHAHAHA

JP 16.03.2016 18:03
Tenho ideia que em certas zonas de portugal também se come qualquer coisa parecida com o estômago.... já vi uma colega minha do trabalho levar... e a minha reação é nhecaaaaa

M.J. 16.03.2016 18:04
Maaaaaaaaaaaaaaaaaaaaaaaaaaggggggggggda! Anda cá que estão a falar dos teus maranhos, acho!

JP 16.03.2016 18:05
Precisamente Image

Maria das Palavras 16.03.2016"

in <https://aprenderumacoisanovapordia.blogspot.pt/vamos-todos-almocar-78127>

24. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

25. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P4 - ID/029-1: Com o tempo, tudo se cura.

Por favor leia o excerto.

(...) -"poucos

Quantas vezes já te partiram o coração?

No meu caso foram duas (como na maioria das mulheres aqui :'), mas a última decepção já foi ultrapassada e desde que resolvi tirar umas férias das mulheres ando melhor ;D mas isso não quer dizer que me vá "virar" para os homens (até porque não sou bissexual) ;)

Quantas vezes já te partiram o coração?

#_COM O TEMPO TUDO SE CURA_#...e que em breve olhes para trás e sintas que foi mesmo ultrapassada!

Quantas vezes já te partiram o coração?

concordo!! até mesmo porque quem vive do passado...é um museu! :devil

Quantas vezes já te partiram o coração?

Ahaha nunca tinha ouvido essa expressão mas sim faz todo o sentido :D

Quantas vezes já te partiram o coração?

vindo de mim faz todo o sentido [smiley=orgulhoso.gif]

Quantas vezes já te partiram o coração?

2 vezes." (...)

in <https://www.rea.pt/forum/index.php?topic=25313.40>

26. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

27. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P4 - ID/029-2: Com o tempo, tudo se cura.

Por favor leia o excerto.

(...) "Agora amiga...cabeça para a frente para a contagem decrescente do grande DIA...que ainda vai ter um sabor mais especial para vocês... Beijinhos aos dois

PARABENS LUCIA E RICARDO

Bom dia o pior ja passou agora e levantar a cabeça e ir em frente para a vossa felicidade se de alguma coisa precisarem estamos aqui obrigado

AMIGA É ASSIM MESMO QUE SE FALA...agora é continuar na azafama dos preparativos e esquecer o que ficou para tras, #_COM O TEMPO TUDO SE CURA_# e tudo se esquece e nós estamos todas aqui tal como no casamento...na alegria e na tristeza, na saude e na doença. Podes contar com a nossa ajuda para tudo o que possas precisar, até para os preparativos se for caso disso as nossas maozinhas estão prontas para entrar em acção. Um beijinho muito grande para os dois e aproveitem muito bem cada dia...Felicidades!

Olá Lucia, só agora acabei de ler tudo o que vos" (...)

in <https://onossocasamento.pt/forum/2892>

28. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

29. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P4 - ID/029-3: Com o tempo, tudo se cura.

Por favor leia o excerto.

(...) “que me definem e que choro com muita facilidade e magoo-me muito facilmente.

A ideia de criar este blog surgiu como um hobbie o meu ex namorado acabou tudo comigo ficamos amigos as coisas entre nós não resultaram davamo-nos mais mal que bem e ele passava a vida toda a tratar-me mal até que eu um dia disse CHEGA e resolvi seguir a minha vida claro que tudo custa ao início e mesmo assim mas #_COM O TEMPO TUDO SE CURA_# e eu sei bem do que falo, ainda não esqueci o meu ex e ainda vai demorar tempo mas eu sei que hei-de conseguir, por isso não se admirem dos meus posts serem todos referentes a esquecer rapazes e coisas do género, mas voltando ao assunto criei um blog porque dá trabalho e exige dedicação e entrega em cada post e sempre me vai ajudando a manter distraída.

Aguardem novidades, beijinhos”

in <https://raparigasecreta.blogs.sapo.pt/2014/09/02/>

30. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

31. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P5 - ID/052-1: O fruto proibido é o mais apetecido.

Por favor leia o excerto.

(...) “em conta o factor genético.

No que se refere à predisposição genética, «o bioquímico Dean Hamer, do Centro Nacional norte-americano do Cancro, detectou uma relação entre a coragem e o gene D4DR. Os portadores de uma versão concreta deste mostram maior inclinação para o risco e o proibido.

Na minha opinião, além da carga genética, este comportamento pode ser explicado tendo em conta o facto de vivermos numa sociedade cheia de regras e normas de bom comportamento, algumas pessoas sentem-se de tal forma prisioneiras destas leis o que as leva a obter prazer transgredindo algo pelo qual seriam repreendidas. Não é o facto de sair de um café sem pagar o que consumiu que dá prazer a essa pessoa, mas sim saber que está a transgredir uma regra e que esse comportamento pode ser punido caso seja descoberto.

Mas como diz o ditado: «#_O FRUTO PROIBIDO É O MAIS APETECIDO_#».”

in https://esquina_da_psicologia.blogs.sapo.pt/759.html

32. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

33. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P5 - ID/052-2: O fruto proibido é o mais apetecido.

Por favor leia o excerto.

(...) “Congresso de Medicina Popular, este ano na sua 26ª edição, está relacionado com o tema do oculto e do místico, condenado pela Igreja e pela ciência, avançou à Lusa o padre Fontes.

Texto da Agência Lusa, editado por Patrícia Viegas
29 Agosto 2012 — 10:01

Facebook
Twitter
Partilhar

Tópicos

Portugal

Relacionados

portugal
Congresso quer evitar descredibilização

O organizador do Congresso de Medicina Popular, em Vilar de Perdizes, Montalegre, padre António Fontes, explicou, recorrendo ao ditado popular “#_O FRUTO PROIBIDO É O MAIS APETECIDO_#”, a motivação das pessoas, vindas de Norte a Sul do país, em visitar o evento.

A 26ª edição do Congresso de Medicina Popular realiza-se de 30 de agosto a 02 de setembro e reúne “profissionais” do oculto, ervanários, curandeiros, exorcistas, médiuns, cartomantes e endreitas para discutir os problemas que inquietam os portugueses.

O padre Fontes, figura emblemática do Barroso conhecido também pela organização da sexta-feira 13, frisou hoje, em conversa com a agência Lusa, que”

in <https://www.dn.pt/portugal/interior/padre-fontes-fruto-proibido-e-o-mais-apetecido--2740945.html>

34. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

35. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P5 - ID/052-3: O fruto proibido é o mais apetecido.

Por favor leia o excerto.

(...) "Porque o fruto proibido...

Existe aquela frase muito conhecida « #_O FRUTO PROIBIDO É O MAIS APETECIDO_# » se concordo com essa frase? Claro! É que concordo mesmo.

Por isso estava a pensar ir à nutricionista e pedir-lhe para me dizer para fazer aquilo que não devo fazer... mas faze-lo e vice-versa.

Ou seja:

Nutricionista - " Sara, você deve comer comidas carregadas de natas, fritos, muita batata, muitas guloseimas carregadas de açúcar e tudo mais. Só não pode comer saladas, grelhados e tudo o que é saudável!"

Visto que o que é saudável seria o tal fruto proibido, podia ser que deixa-se de comer as porcarias todas e começasse a comer o saudável... não sei, digo eu. Visto que eu sou completamente do contra se calhar até funcionava.

(Mas eu até adoro comida saudável... mas também não resisto a umas belas natas e estraga tudo... malditas!!!)

É como uma vez na creche a educadora para um miúdo: " Não mexas na televisão!" e ele mexe " Gonçalo! Eu disse para não mexeres na televisão" e ele continua lá com a mão " Está bem" (...)

in <https://sarainwonderland.blogs.sapo.pt/123352.html>

36. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

37. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P6 - ID/059-1: O sol quando nasce é para todos.

Por favor leia o excerto.

(...) “contrasta com a produção hídrica e eólica que asseguram a maioria da produção verde. Na lógica da transição energética, esta aposta tem de acontecer sempre com o foco de atingir o consumo suficiente de energia, pois qualquer produção de energia tem implicações várias na sustentabilidade e nem todas são positivas, mesmo tratando-se de energias renováveis.

Para além da sustentabilidade inerente, as energias renováveis são também relevantes pela forma democrática como podem ser aproveitadas e utilizadas. #_O SOL QUANDO NASCE É PARA TODOS_# e deve ser um desígnio de Portugal apostar na democratização do acesso ao mesmo. A promoção do autoconsumo, a nível individual e coletivo (por exemplo, nos condomínios, em bairros e comunidades) é um passo fundamental que Portugal tarda em dar, ao contrário do que outros países já estão a fazer. Na União Europeia já se percebeu que existe um enorme potencial de fomento de sistemas descentralizados de produção de energia renovável, atribuindo a cada cidadão” (...)

in <https://observador.pt/opiniao/o-sol-quando-nasce-e-para-todos/>

38. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

39. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P6 - ID/059-2: O sol quando nasce é para todos.

Por favor leia o excerto.

(...) "que integram o projecto circularam pela escola e surpreenderam professores, funcionários e restantes colegas oferecendo-lhes uma flor. Um modo original de dar as boas-vindas à Primavera. "Uma iniciativa diferente e que nos motiva a ser melhores", aponta Carina que sonha em ser actriz. Ao seu lado Filipa Fernandes, a técnica de intervenção local que os acompanha permanentemente no terreno, sorri de orgulho com uma flor amarela. Celeste Simão, técnica disciplinar na equipa móvel multidisciplinar do Programa para Prevenção e Eliminação da Exploração do Trabalho Infantil (PETI), explica que a iniciativa pretendeu integrar ainda mais estes alunos no seio da comunidade escolar, onde porém, já são bem aceites por todos. O PIEF da Chamusca acolhe jovens dos concelhos de Torres Novas, Entroncamento, Vila Nova da Barquinha e Golegã. Neste momento, os alunos do PIEF encontram-se a construir a sua mascote, um moinho. As actividades encontram-se descritas em <http://piefdachamusca.blogspot.com>. O PIEF é uma medida de excepção criada pelo governo que representa uma alternativa de vida para os jovens quando tudo o resto falhou. Porque #_O SOL QUANDO NASCE É PARA TODOS_#. Elsa Ribeiro Gonçalves" (...)

in <http://omirante.pt/semanario/2009-03-26/primeiro-plano/2009-03-25-o-sol-quando-nasce-e-para-todos>

42. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

43. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P7 - ID/077-1: Quem com ferro fere com ferro será ferido.

Por favor leia o excerto.

(...) “o segundo dia de campanha oficial da coligação Portugal à Frente dá vontade de dizer que #_“QUEM COM FERRO FERRE, COM FERRO SERÁ FERIDO”._# Passos Coelho - que tanto tem gasto palavras nas suas intervenções a sublinhar as suas "contas certas" e acusando António Costa de esconder as do próprio programa (em matéria de cortes no sistema de segurança social) - protagonizou um "equivoco" de milhões de euros.

Na sua intervenção, ontem, em Beja, Pedro Passos Coelho anunciou que no próximo dia 15 de outubro o seu governo iria pagar "mais 5400 milhões de euros dos 78 milhões" de ajuda externa concedida pelo Fundo Monetário Internacional (FMI), Comissão Europeia e Banco Central Europeu, em 2011. Depois de um almoço-comício regado com um oportuno vinho alentejano de nome Amnésia, Passos baralhou-se. Afinal, como viria a emendar à noite, o ajuste de contas era, nesse mesmo valor, na mesma data, mas em "obrigações do Tesouro contraídas há 10 anos pelo governo socialista". "Quando tenho lapsos, corrijo-os sem qualquer problema", reconheceu.

Num dia que começou, como os anteriores, com uma visita a uma grande herdade" (...)

in <https://www.dn.pt/politica/interior/clientelas-equivocos-e-um-vinho-amnesia-para-passos-4791146.html>

44. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

45. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P7 - ID/077-2: Quem com ferro fere com ferro será ferido.

Por favor leia o excerto.

(...) “causar muitas desavenças sociais.

Muito embora somos parte de um só planeta e vamos todos em direção do mesmo destino, independente se um estudou em Haward e outro sequer sabe as vogais. No fim, todos vamos encerrar nossas carreiras e deixar nossos herdeiros – quer de sangue, quer de patrimônios – seguindo nosso destino, perpetuando-nos até o fim dos tempos, porém, todos vamos sucumbir ao desligamento e avohai.

Hoje eu fui chamado para refletir sobre a frase: #_“QUEM COM FERRO FERRE, COM FERRO SERÁ FERIDO.”_# A frase serviu para mim. A carapuça me serviu como uma luva e eu aprendi a lição da Lei do Retorno.

A Lei do Retorno acontece gratuitamente, pois tudo o que você dá de bom coração, volta de bom coração. Se der com coração leviano, você está no sal. Se não der, não receberá. Tudo é um toma lá e dá cá, não há espaço para refutação nem questionamento. É pá, pum! Sem erro.

Então” (...)

In <https://blogdorodrigocaldeira.blogs.sapo.pt/38112.html>

46. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

47. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P7 - ID/077-3: Quem com ferro fere com ferro será ferido.

Por favor leia o excerto.

(...) “a admitir a conexão que existe entre eles. O homem comum, nos seus interesses mesquinhos, não considera a dor senão como resgate e pagamento, desconhecendo o gozo de padecer por cooperar, sinceramente, na edificação do Reino do Cristo. Aquele que se compraz na caminhada pelos atalhos do mal, a própria Lei se incumbirá de trazê-lo de retorno às vias do bem. O passado, muitas vezes, determina o presente que, por sua vez, determina o futuro. #_“QUEM COM FERRO FERRE, COM FERRO SERÁ FERIDO”_# (7), disse o Mestre. Porém, cabe uma ressalva, nem todo sofrimento é expiação.

No item 9, cap. V, de O Evangelho Segundo o Espiritismo, Allan Kardec assinala: “Não se deve crer, entretanto, que todo sofrimento porque se passa neste mundo seja, necessariamente, o indício de uma determinada falta: trata-se, freqüentemente, de simples provas escolhidas pelo Espírito para sua purificação, para acelerar o seu adiantamento”.(8).

São claras as palavras do Codificador. Não estão corretos aqueles” (...)

In <https://jehannedarc.blogs.sapo.pt/23111.html>

48. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

49. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P8 - ID/018-1: A vingança serve-se fria.

Por favor leia o excerto.

(...) “vez menos bem-vindas. O líder húngaro ameaça com a constituição de uma “grande coligação” europeia dos partidos que pensam como ele em relação aos imigrantes e aos refugiados, que poderia alterar radicalmente o actual quadro político no Parlamento Europeu. Seria ou será o prenúncio de um novo mapa político europeu que não deixa ninguém tranquilo.

4. Entretanto, já pouca gente se dá ao trabalho de esconder um certo prazer de ver a chanceler em risco. #_A VINGANÇA SERVE-SE FRIA_# e chegou finalmente a hora de confrontar a chanceler com alguns dos seus erros europeus, desde o “momento hegemónico” ou “unilateral” da Alemanha, durante boa parte da crise do euro, até à sua decisão de abrir as portas aos refugiados da guerra na Síria sem se dar ao trabalho de consultar os parceiros. A crítica é verdadeira. O poder traz sempre consigo mais responsabilidade que nem sempre a chanceler quis assumir. Mas talvez seja má” (...)

In <https://www.publico.pt/2018/07/08/mundo/opiniao/a-vinganca-servese-fria-1837258>

50. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

51. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P8 - ID/018-2: A vingança serve-se fria.

Por favor leia o excerto.

(...) “até 2012 para fazer este vinho mágico. Pretendendo homenagear Dona Antónia Adelaide Ferreira, Luís Sottomayor deu ao mundo um vinho de cor dourada, com aromas a melão e a pêra, notas florais e minerais, demonstrando enorme volume na boca, mas também acidez e frescura, e madeira bem integrada apesar do longo estágio de 12 meses em barrica de carvalho francês.

Quando desafiámos o vinho a defrontar-se ao lado de uma açorda de lavagante com filete de peixe galo, chegámos a temer o pior, mas ... casaram e foram felizes para sempre.

Como este (mais um adjectivo) precioso néctar está limitado a 1000 garrafas, urge disputá-las, uma a uma. Pela minha parte, irei à tasca à procura do meu interlocutor, mudarei o rótulo, colocarei uma tampa de plástico e perguntarei, desafiador: “Boa pinga, há?”. #_A VINGANÇA SERVE-SE FRIA_#, entre os 10 e os 15°C.

Casa Ferreirinha Antónia - Adelaide Ferreira Branco

Ano: 2012

Região: Douro”

In <https://sol.sapo.pt/artigo/124434/a-vinganca-serve-se-fria->

52. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

53. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P8 - ID/018-3: A vingança serve-se fria.

Por favor leia o excerto.

“Vingança, onde a comida tem três temperaturas
21.08.2017 13:00 por Ana Maria Henriques 0

No Vingança, a carta está dividida em três e o elemento mais importante é a temperatura da comida servida

Vamos arrumar com a frase feita logo no início deste texto: aqui, #_ A VINGANÇA SERVE-SE FRIA_#, morna e quente, num conceito de restaurante explicado aos clientes assim que se sentam à mesa. No Vingança, o espaço que é irmão do Cruel, na agora repleta de restaurantes Rua da Picaria, a carta está dividida em três e o elemento mais importante é a temperatura da comida servida.

Logo a começar, um couvert (€1,50/pessoa) para demonstrar o conceito tripartido: manteiga fria com flor de sal, tomate assado servido morno e creme de ervilhas com azeite de chouriço. Destaque para o tomate assado, que se desfaz quando acompanhado por uma fatia de pão.

Tal como no Cruel - cuja carta também é assinada pela” (...)

In <https://www.sabado.pt/gps/gourmet/restaurantes/detalhe/vinganca-onde-a-comida-tem-tres-temperaturas>

54. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

55. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P9 - ID/080-1: Quem está vivo sempre aparece.

Por favor leia o excerto.

"#_QUEM ESTÁ VIVO, SEMPRE APARECE_#. Ora viva a greve e esta em particular, a greve geral do dia 24 de Novembro. De facto, a greve não é filha de um Deus menor a necessitar de "sindicância" com números fiéis, amigos caninos, que existem como existem. Se a questão se esgotasse no descobrir o Wally na verdade verdadeira desta iniciativa, ora, ora talvez fosse do interesse nacional a criação de um fantástico protocolo, muito, mas mesmo muito aplicável, com manual de instruções em português, claro, e já agora aprovado de livre e espontânea vontade por todas as partes porque, é bom não esquecer, no caso dos trabalhadores do sector público, o governo e o patrão acumulam funções no mais perfeito sincronismo.

Ao que interessa: o que antecede a greve é o que a justifica: o reconhecimento pelos próprios sindicatos de que toda a argumentação e iniciativas falharam. E é desta falha, da perda que se tem" (...)

In <https://jugular.blogs.sapo.pt/3014537.html>

56. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

57. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P9 - ID/080-2: Quem está vivo sempre aparece.

Por favor leia o excerto.

(...) “interromper duas vezes por causa do trabalho, razão pela qual só falarei mais detalhadamente sobre ele quando conseguir chegar ao fim. Hoje é só mesmo para levantar o dedo e dizer que as férias foram boas e vi, felizmente, muita gente a ler – mães e tudo, que são as que têm sempre os olhos nos miúdos e, por isso, quase nunca podem estender-se descansadas com o seu livrinho; claro que a maioria se ocupava com o best-seller pretensamente pornográfico As Cinquenta Sombras de Gray em várias línguas, mas também havia quem se dedicasse a biografias, ensaios e romances mais suculentos. Voltar ao trabalho e às leituras obrigatórias é que foi difícil, mas hei-de sobreviver. Se estiverem aí desse lado, vai ser, garanto, muito mais fácil. Obrigada por não se terem esquecido deste blogue.

Bom dia! Bons olhos a vejam! Ou: #_QUEM ESTÁ VIVO SEMPRE APARECE_#.

Comecei a ler Mau Tempo no Canal, de Vitorino Nemésio.”

In <https://horasextraordinarias.blogs.sapo.pt/138186.html>

58. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

59. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P9 - ID/080-3: Quem está vivo sempre aparece.

Por favor leia o excerto.

(...) “Ah que belo Dia de Reis!
d'oliveira

Minha Querida Kamikase,

Eu já tenho poucas coisas realmente interessantes em que acreditar ou, como dizia o outro, sou macaco de rabo pelado. Contudo, por estouvamento consubstancial, por ingenuidade suicidária ou por burrice congénita sempre acreditei nalgumas máximas que o nosso povo, lingüareiro e facilitador, usa e abusa. Uma delas será: #_QUEM ESTÁ VIVO SEMPRE APARECE_# ou estoutra: não há mal que sempre dure...

Tudo isto para Lhe dizer, aqui e não em comentário ao seu texto (está cheio de caracteres horríveis e inidentificáveis...) da grande alegria em vê-la regressar, triunfal e acutilante, mais "vento divino" do que nunca, e a arrear a giga comme il faut! Já telefonei (logo eu tão inimigo desse objecto fatal...) a não sei quantas pessoas para acudirem a este jovial ajuntamento e lerem com os próprios olhos como se destila meio quilo de cicuta em estado puro.

Eu nunca fui muito de” (...)

In <https://incursoes.blogs.sapo.pt/766438.html?thread=2377190>

60. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

61. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P10 - ID/039-1: Gato escaldado, de água fria tem medo.

Por favor leia o excerto.

(...) “Novo Banco e a “dívida senior”. Gato escaldado de água fria tem medo, dizem os analistas

23/2/2016, 23:08

Obrigações "senior" que o Banco de Portugal protegeu estão a negociar a 70% do valor. Instituição garantiu que não haverá mais perdas, mas especialistas duvidam. Há uma proposta para "corrigir o erro"

Once bitten, twice shy ou, em português, #_GATO ESCALDADO DE ÁGUA FRIA TEM MEDO_#.

A negociação em mercado dos títulos de dívida que foram protegidos na decisão polémica do Banco de Portugal mostra que os investidores continuam muito receosos em relação ao futuro do Novo Banco. Apesar de o Banco de Portugal garantir que a resolução está “completada” e que não haverá mais perdas, as obrigações do Novo Banco estão a negociar a pouco mais de 70% do seu valor e o risco está ao nível da Grécia, pese embora o facto de o banco ter agora uma solvabilidade mais confortável. Este é um sinal claro de que os investidores duvidam das garantias dadas pela instituição liderada por Carlos Costa e admitem que as perdas podem não ter ficado por aqui.”

In <https://observador.pt/2016/02/23/novo-banco-divida-senior-gato-escaldado-agua-fria-medo-dizem-os-analistas/>

62. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

63. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P10 - ID/039-2: Gato escaldado, de água fria tem medo.

Por favor leia o excerto.

“GATO ESCALDADO DE ÁGUA FRIA TEM MEDO.

A moção de estratégia ao congresso do CDS-PP, que tem como primeiro subscritor Pires de Lima, contém as linhas mestras do programa eleitoral do partido de Paulo Portas às próximas legislativas, sejam em 2014 ou em 2015: reposição progressiva dos rendimentos de funcionários públicos e pensionistas, desagravamento progressivo do IRS e da sobretaxa, aumento do salário mínimo nacional. Esta moção da direcção do CDS-PP está em linha com a sua estratégia bipolar: estar no governo e na oposição, ao mesmo tempo. Espezinhar funcionários públicos e pensionistas nas decisões do conselho de ministros e no parlamento e dizer que está contra tais medidas na sede do partido e nos corredores ministeriais. Ao mesmo tempo, pisca o olho ao PS (fim da austeridade, retoma da economia e do emprego). #_COMO GATO ESCALDADO DE ÁGUA FRIA TEM MEDO_#, pode ser que lhes saia o tiro pela culatra. Aguardemos.”

In <https://hojehaconquilhas.blogs.sapo.pt/1334626.html>

64. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

65. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

P10 - ID/039-3: Gato escaldado, de água fria tem medo.

Por favor leia o excerto.

“Gato escaldado

Em resposta à proposta do presidente da Junta de Freguesia de Alhandra para que seja retirada a sebe à volta do fontanário da praça 7 de Março, a presidente da Câmara de Vila Franca indicou que a população seja auscultada sobre o assunto. Depois da polémica com o monumento à tauromaquia, Maria da Luz Rosinha não quer arriscar.
#_GATO ESCALDADO DE ÁGUA FRIA TEM MEDO_#!”

In <http://omirante.pt/semanario/2006-08-09/guarda-rios/2006-08-08-gato-escaldado>

66. Coloque aqui a sua resposta (copie e cole para o campo abaixo o elemento ou a expressão do texto que considerar mais relevante): *

67. O que o provérbio diz sobre esses elementos é: *

Marcar apenas uma oval.

- positivo
- mais positivo do que negativo
- nem positivo nem negativo
- mais negativo do que positivo
- negativo
- não sei o que responder
- não é possível, com base no excerto apresentado, determinar a que se refere o provérbio

Terminou o seu questionário.

Muito obrigada pela sua participação!

Com tecnologia



.6 Mínimo paremiológico do português europeu

Nº Provérbio

- 1 A beleza está nos olhos de quem a vê.
- 2 A carne é fraca.
- 3 A cavalo dado não se olha ao dente.
- 4 A culpa morreu solteira.
- 5 A curiosidade matou o gato.
- 6 A descer todos os santos ajudam.
- 7 A esperança é sempre a última a morrer.
- 8 A experiência é a mãe da ciência.
- 9 A falar é que a gente se entende.
- 10 A fé move montanhas.
- 11 A fome é o melhor tempero.
- 12 A galinha da vizinha é sempre melhor que a minha.
- 13 A idade não perdoa.
- 14 A intenção é que conta.
- 15 A justiça tarda mas não falta.
- 16 A laranja de manhã é ouro, à tarde é prata e à noite mata.
- 17 A mentira tem perna curta.
- 18 A montanha pariu um rato.
- 19 A morte não escolhe idades.
- 20 A mulher e a sardinha querem-se da mais pequenina.
- 21 A noite é boa conselheira.
- 22 À noite todos os gatos são pardos.
- 23 A ocasião faz o ladrão.
- 24 A paciência tem limites.
- 25 A pensar morreu um burro.
- 26 A preguiça é a mãe de todos os vícios.
- 27 A pressa é inimiga da perfeição.
- 28 A roupa suja lava-se em casa.
- 29 A saúde não tem preço.
- 30 A união faz a força.
- 31 A verdade vem sempre ao de cima.
- 32 A vida são dois dias.
- 33 A vingança serve-se fria.
- 34 A voz do povo é a voz de Deus.
- 35 Abril, águas mil.
- 36 Achado não é roubado.
- 37 Água mole em pedra dura tanto bate até que fura.
- 38 Águas passadas não movem moinhos.
- 39 Albarde-se o burro à vontade do dono.
- 40 Amanhã é outro dia.
- 41 Amanhã também é dia.
- 42 Amigo não empata amigo.

- 43 Amigos, amigos, negócios à parte.
- 44 Amor com amor se paga.
- 45 Ano novo, vida nova.
- 46 Antes só que mal acompanhado.
- 47 Aprender até morrer.
- 48 As ações ficam com quem as pratica.
- 49 As aparências iludem.
- 50 As árvores morrem de pé.
- 51 As más notícias chegam depressa.
- 52 As paredes têm ouvidos.
- 53 Até ao lavar dos cestos é vindima.
- 54 Azar ao jogo, sorte no amor.
- 55 Beleza não põe mesa.
- 56 Boas contas fazem os bons amigos.
- 57 Boda molhada, boda abençoada.
- 58 Burro velho não aprende línguas.
- 59 Cá se fazem, cá se pagam.
- 60 Cada cabeça sua sentença.
- 61 Cada cavadela sua minhoca.
- 62 Cada coisa a seu tempo.
- 63 Cada macaco no seu galho.
- 64 Cada maluco com a sua mania.
- 65 Cada qual com o seu igual.
- 66 Cada um é como cada qual.
- 67 Cada um é para o que nasce.
- 68 Cada um por si e Deus por todos.
- 69 Cada um puxa a brasa à sua sardinha.
- 70 Cada um sabe as linhas com que se cose.
- 71 Cada um tem aquilo que merece.
- 72 Candeia que vai à frente alumia duas vezes.
- 73 Cão que ladra não morde.
- 74 Casa onde não há pão, todos ralham e ninguém tem razão.
- 75 Casa roubada, trancas à porta.
- 76 Cautela e caldos de galinha nunca fizeram mal a ninguém.
- 77 Com o fogo não se brinca.
- 78 Com papas e bolos se enganam os tolos.
- 79 Com uma mentira se apanha uma verdade.
- 80 Contas são contas.
- 81 Contra factos não há argumentos.
- 82 Cresce e aparece.
- 83 Da discussão nasce a luz.
- 84 De boas intenções está o inferno cheio.
- 85 De Espanha nem bom vento, nem bom casamento.
- 86 De pequenino é que se torce o pepino.
- 87 De poeta e de louco, todos temos um pouco.
- 88 Deitar cedo e cedo erguer, dá saúde e faz crescer.
- 89 Depois da tempestade vem a bonança.
- 90 Depressa e bem não há quem.

- 91 Deus dá nozes a quem não tem dentes.
- 92 Deus escreve direito por linhas tortas.
- 93 Deus não dorme.
- 94 Devagar se vai ao longe.
- 95 Dia de São Martinho, lume, castanhas e vinho.
- 96 Diz-me com quem andas, dir-te-ei quem és.
- 97 Do mal, o menos.
- 98 Dos fracos não reza a História.
- 99 É mais fácil apanhar um mentiroso que um coxo.
- 100 É pior a emenda que o soneto.
- 101 É preciso dar tempo ao tempo.
- 102 Em boca fechada não entra mosca.
- 103 Em casa de ferreiro, espeto de pau.
- 104 Em Roma sê romano.
- 105 Em tempo de guerra não se limpam armas.
- 106 Em terra de cegos quem tem olho é rei.
- 107 Enquanto há vida há esperança.
- 108 Enquanto o pau vai e vem folgam as costas.
- 109 Entre marido e mulher, nunca metas a colher.
- 110 Entre mortos e feridos alguém há-de escapar.
- 111 Errar é humano.
- 112 Faz o que eu digo e não faças o que eu faço.
- 113 Filho de peixe sabe nadar.
- 114 Filhos criados, trabalhos dobrados.
- 115 Gaivotas em terra, tempestade no mar.
- 116 Gato escaldado de água fria tem medo.
- 117 Gordura é formosura.
- 118 Grande nau, grande tormenta.
- 119 Grão a grão enche a galinha o papo.
- 120 Há gente para tudo.
- 121 Há mais marés que marinheiros.
- 122 Há males que vêm por bem.
- 123 Há mar e mar, há ir e voltar.
- 124 Há sempre uma primeira vez para tudo.
- 125 Hoje por mim, amanhã por ti.
- 126 Ladrão que rouba ladrão tem cem anos de perdão.
- 127 Longe dos olhos, longe do coração.
- 128 Lua de mel, lua de fel.
- 129 Mãe há só uma.
- 130 Mais vale cair em graça do que ser engraçado.
- 131 Mais vale pouco que nada.
- 132 Mais vale prevenir do que remediar.
- 133 Mais vale tarde do que nunca.
- 134 Mais vale um pássaro na mão do que dois a voar.
- 135 Manda quem pode.
- 136 Mãos frias, coração quente.
- 137 Melhor é dobrar que quebrar.
- 138 Muita parra, pouca uva.

- 139 Muito riso, pouco siso.
140 Na hora do aperto é que se conhecem os amigos.
141 Não deixes para amanhã o que podes fazer hoje.
142 Não é com vinagre que se apanham moscas.
143 Não faças aos outros o que não queres que te façam a ti.
144 Não há amor como o primeiro.
145 Não há bela sem senão.
146 Não há bem que sempre dure nem mal que nunca acabe.
147 Não há duas sem três.
148 Não há regra sem exceção.
149 Não há rosa sem espinhos.
150 Não se fazem omeletes sem ovos.
151 Não se pode ter tudo.
152 Nem oito nem oitenta.
153 Nem só de pão vive o homem.
154 Nem tanto ao mar, nem tanto à terra.
155 Nem tudo o que reluz é ouro.
156 Nem tudo o que vem à rede é peixe.
157 Ninguém é de ferro.
158 Ninguém nasce ensinado.
159 No meio é que está a virtude.
160 No melhor pano cai a nódoa.
161 No poupar é que está o ganho.
162 Nunca digas nunca.
163 Nunca é tarde para aprender.
164 O amor é cego.
165 O ataque é a melhor defesa.
166 O barato sai caro.
167 O bom filho à casa torna.
168 O crime não compensa.
169 O dinheiro não traz felicidade.
170 O fruto proibido é o mais apetecido.
171 O futuro a Deus pertence.
172 O hábito não faz o monge.
173 O meu é meu, o teu é nosso.
174 O pior cego é o que não quer ver.
175 O primeiro milho é dos pardais.
176 O prometido é devido.
177 O que arde cura.
178 O que é bom acaba depressa.
179 O que é doce nunca amargou.
180 O que lá vai lá vai.
181 O que não mata, engorda.
182 O que não tem remédio, remediado está.
183 O que tem de ser tem muita força.
184 O saber não ocupa lugar.
185 O segredo é a alma do negócio.
186 O seguro morreu de velho.

- 187 O seu a seu dono.
188 O silêncio é de ouro.
189 O sol quando nasce é para todos.
190 O tempo é dinheiro.
191 O tempo tudo cura.
192 O tempo voa.
193 Olho por olho, dente por dente.
194 Olhos que não veem, coração que não sente.
195 Onde há fumo, há fogo.
196 Os amigos são para as ocasiões.
197 Os cães ladram e a caravana passa.
198 Os erros pagam-se caro.
199 Os fins justificam os meios.
200 Os gostos não se discutem.
201 Os homens não se medem aos palmos.
202 Os olhos comem mais que a barriga.
203 Os olhos são o espelho da alma.
204 Os opostos atraem-se.
205 Os últimos serão os primeiros.
206 Ou oito ou oitenta.
207 Palavra puxa palavra.
208 Palavras loucas, orelhas moucas.
209 Palavras, leva-as o vento.
210 Pão pão, queijo queijo.
211 Para a frente é que é o caminho.
212 Para bom entendedor, meia palavra basta.
213 Para grandes males, grandes remédios.
214 Para morrer basta estar vivo.
215 Para quem é, bacalhau basta.
216 Parar é morrer.
217 Patrão fora, dia santo na loja.
218 Peixe não puxa carroça.
219 Pela boca morre o peixe.
220 Perdido por cem, perdido por mil.
221 Por causa de uns, pagam os outros.
222 Por morrer uma andorinha não acaba a Primavera.
223 Porta da rua é serventia da casa.
224 Quando a cabeça não tem juízo, o corpo é que paga.
225 Quando a esmola é grande, o pobre desconfia.
226 Quando o mar bate na rocha quem se lixa é o mexilhão.
227 Quando um burro fala, os outros baixam as orelhas.
228 Quando uma porta se fecha outra se abre.
229 Quanto mais alto se sobe, maior é a queda.
230 Quanto mais depressa mais devagar.
231 Quanto mais se tem, mais se quer.
232 Quem anda à chuva molha-se.
233 Quem boa cama fizer nela se deitará.
234 Quem brinca com o fogo queima-se.

- 235 Quem cala consente.
236 Quem canta seus males espanta.
237 Quem casa quer casa.
238 Quem com ferro fere com ferro será ferido.
239 Quem conta um conto acrescenta um ponto.
240 Quem corre por gosto não cansa.
241 Quem dá aos pobres empresta a Deus.
242 Quem dá e torna a tirar, ao inferno vai parar.
243 Quem desdenha quer comprar.
244 Quem diz a verdade não merece castigo.
245 Quem diz o que quer, ouve o que não quer.
246 Quem é vivo sempre aparece.
247 Quem espera sempre alcança.
248 Quem espera, desespera.
249 Quem está mal que se mude.
250 Quem estraga velho paga novo.
251 Quem fala assim não é gago.
252 Quem faz um cesto faz um cento.
253 Quem feio ama bonito lhe parece.
254 Quem foi ao mar, perdeu o lugar.
255 Quem mais jura mais mente.
256 Quem muito dorme pouco aprende.
257 Quem muito escolhe pouco acerta.
258 Quem muito fala, pouco acerta.
259 Quem não arrisca não petisca.
260 Quem não chora não mama.
261 Quem não deve não teme.
262 Quem não é por mim é contra mim.
263 Quem não é visto não é lembrado.
264 Quem não quer ser lobo não lhe veste a pele.
265 Quem não sabe é como quem não vê.
266 Quem não tem cão caça com gato.
267 Quem não tem dinheiro não tem vícios.
268 Quem não trabuca não manduca.
269 Quem nasce torto tarde ou nunca se endireita.
270 Quem pergunta quer saber.
271 Quem pode pode.
272 Quem procura acha.
273 Quem ri por último ri melhor.
274 Quem sabe sabe.
275 Quem sai aos seus não degenera.
276 Quem se mete em atalhos, mete-se em trabalhos.
277 Quem semeia ventos colhe tempestades.
278 Quem semeia, colhe.
279 Quem te avisa teu amigo é.
280 Quem tem boca vai a Roma.
281 Quem tem cu tem medo.
282 Quem tem medo compra um cão.

- 283 Quem tem telhados de vidro não atira pedras ao do vizinho.
284 Quem tem unhas é que toca guitarra.
285 Quem tudo quer tudo perde.
286 Quem vai à guerra dá e leva.
287 Quem vai para o mar avia-se primeiro em terra.
288 Quem vê caras não vê corações.
289 Quem vier atrás que feche a porta.
290 Querer é poder.
291 Recordar é viver.
292 Rei morto, rei posto.
293 Rir é o melhor remédio.
294 Roma e Pavia não se fizeram num dia.
295 Saber esperar é uma virtude.
296 Santos da casa não fazem milagres.
297 São mais as vozes que as nozes.
298 Se Maomé não vai à montanha, a montanha vai a Maomé.
299 Sem trabalho nada se consegue.
300 Só se vive uma vez.
301 Tal pai, tal filho.
302 Todos os caminhos vão dar a Roma.
303 Toma lá, dá cá.
304 Três foi a conta que Deus fez.
305 Tristezas não pagam dívidas.
306 Tudo está bem quando acaba bem.
307 Um abismo chama outro abismo.
308 Um burro carregado de livros é um doutor.
309 Um crime não justifica outro.
310 Um homem prevenido vale por dois.
311 Um mal nunca vem só.
312 Uma imagem vale mais que mil palavras.
313 Uma mão lava a outra.
314 Vão-se os anéis, ficam os dedos.
315 Velhos são os trapos.
316 Vivendo e aprendendo.
317 Vozes de burro não chegam ao céu.
318 Zangam-se as comadres, descobrem-se as verdades.

.7 100 jogos com provérbios com disponibilidade lexical

Tabela 3: 100 jogos com provérbios com disponibilidade lexical

ID	ID-Q2	Provérbio	Tipo de jogo	Provérbio e resposta	RC	DL-A	DL-MP
LP_SP00772	[Q-001]	A descer todos os santos ajudam.	Selecione a palavra em falta no provérbio.	A descer todos os ... ajudam. santos	x	2	2
				anjos			
				diabos			
				demônios			
LPP_JRMC30111	[Q-002]	Depois da tempestade vem a bonança.	Selecione a palavra em falta no provérbio.	Depois da tempestade vem a ... bonança	x	2	2
				felicidade			
				tranquilidade			
				esperança			
LPP_JRMC02817	[Q-003]	Os homens não se medem aos palmos.	Selecione a palavra em falta no provérbio.	Os ... não se medem aos palmos. homens	x	2	2
				cavaleiros			
				jovens			
				líderes			
LP_SP34192	[Q-004]	Quem não tem dinheiro não tem vícios.	Selecione a palavra em falta no provérbio.	Quem não tem ... não tem vícios. dinheiro	x	2	2
				fortuna			
				benefícios			
				ofícios			
LP_SP01151	[Q-005]	A galinha da vizinha é sempre melhor que a minha.	Selecione a palavra em falta no provérbio.	A ... da vizinha é sempre melhor que a minha. galinha	x	2	2
				vaca	x		
				burrinha			
				andorinha			
PP_AM11522	[Q-006]	São mais as vezes que as nozes.	Selecione a palavra em falta no provérbio.	São mais as vezes que as ... nozes	x	2	2
				doses			
				cantigas			
				marimbas			
LP_SP18387	[Q-007]	Mais vale tarde do que nunca.	Selecione a palavra em falta no provérbio.	Mais vale tarde do que ... nunca	x	2	2
				sempre			
				cedo			
				jamais			
LPP_JRMC28402	[Q-008]	O tempo voa.	Selecione a palavra em falta no provérbio.	O ... voa. tempo	x	2	2
				dinheiro	x		
				vento			
				cheiro			
LP_SP36698	[Q-009]	Santos de casa não fazem milagres.	Selecione a palavra em falta no provérbio.	Santos de casa não fazem ... milagres	x	2	2
				padroeiros			

Tabela 3 continuação da página anterior

ID	ID-Q2	Provérbio	Tipo de jogo	Provérbio e resposta	RC	DL-A	DL-MP
LPP_JRMC10969	[Q-010]	Melhor é dobrar que quebrar.	Seleção a palavra em falta no provérbio.	maravilhas prodígios Melhor é dobrar que... quebrar morrer abalar render	x	2	2
LPP_JRMC05525	[Q-011]	Quem tem unhas é que toca guitarra.	Seleção a palavra em falta no provérbio.	unhas garra palheta cunhas	x	1	2
LP_SP09135	[Q-012]	Com o tempo, tudo se cura.	Seleção a palavra em falta no provérbio.	Com o ..., tudo se cura. tempo futuro amanhã momento	x	2	2
LPP_JRMC08894	[Q-013]	A curiosidade matou o gato.	Seleção a palavra em falta no provérbio.	A curiosidade matou o ... gato rato sapo frade	x	2	2
PP_AM08194	[Q-014]	O que arde cura.	Seleção a palavra em falta no provérbio.	O que arde ... cura dura fura ajuda	x	2	2
LPP_JRMC20526	[Q-015]	Quem está mal que se mude.	Seleção a palavra em falta no provérbio.	Quem está mal que se ... mude grude refugie cale	x	2	2
LP_SP11415	[Q-016]	Deus não dorme.	Seleção a palavra em falta no provérbio.	... não dorme Deus Zeus Mateus semideus	x	2	2
LPP_JRMC03122	[Q-017]	À primeira cavadela, minhoca.	Seleção a palavra em falta no provérbio.	À primeira cavadela ... minhoca agúdias feijoca mandioca	x	2	1
LPP_JRMC00161	[Q-018]	A má acção fica com quem a pratica.	Seleção a palavra em falta no provérbio.	A ... acção fica com quem a pratica. má boa maldosa	x	2	2

Tabela 3 continuação da página anterior

ID	ID-Q2	Provérbio	Tipo de jogo	Provérbio e resposta	RC	DL-A	DL-MP
GLP_JPM26824	[Q-019]	Três foi a conta que Deus fez.	Selecione a palavra em falta no provérbio.	pior Três foi a conta que ... fez.	x	2	2
				Deus Zeus			
				Mateus português			
LPP_JRMC06608	[Q-020]	O pior cego é o que não quer ver.	Selecione a palavra em falta no provérbio.	O pior ... é o que não quer ver. cego	x	2	2
				desconhecedor indouto			
				ignorante			
LP_SP03257	[Q-021]	A vingança serve-se fria.	Selecione a palavra em falta no provérbio.	A vingança serve-se ... fria	x	2	2
				mansa gelada			
				quente			
LP_SP03372	[Q-022]	Abril, águas mil.	Selecione a palavra em falta no provérbio.	Abril, ... mil. águas	x	2	2
				chuvas	x		
				mágoas marés			
PP_AM07400	[Q-023]	Nem tanto ao mar, nem tanto à terra.	Selecione a palavra em falta no provérbio.	Nem tanto ao ... nem tanto à terra. mar	x	2	2
				ar oceano			
				lago			
LPP_JRMC26442	[Q-024]	Não há rosa sem espinhos.	Selecione a palavra em falta no provérbio.	Não há rosa sem ... espinhos	x	2	2
				estames			
				picos botão			
LPP_JRMC31511	[Q-025]	O hábito não faz o monge.	Selecione a palavra em falta no provérbio.	O ... não faz o monge. hábito	x	2	2
				vestuário capote			
				costume			
PP_AM00559	[Q-026]	A noite é boa conselheira.	Selecione a palavra em falta no provérbio.	A ... é boa conselheira. noite	x	2	2
				almofada amiga	x		
				vigília			
PP_AM09826	[Q-027]	Quem anda à chuva molha-se.	Selecione a palavra em falta no provérbio.	Quem anda à chuva ... molha-se	x	2	2
				encharca-se banha-se			
				lava-se			

Tabela 3 continuação da página anterior

ID	ID-Q2	Provérbio	Tipo de jogo	Provérbio e resposta	RC	DL-A	DL-MP
PP_AM00785	[Q-028]	A roupa suja lava-se em casa.	Selecione a palavra em falta no provérbio.	A roupa ... lava-se em casa. suja	x	2	2
				encardida			
				imaculada			
LPP_JRMC23163	[Q-029]	Parar é morrer.	Selecione a palavra em falta no provérbio.	Parar é ...	x	2	2
				morrer			
				perder			
				fraquejar			
				perecer			
GLP_JPM23934	[Q-030]	Quem semeia ventos colhe tempestades.	Selecione a palavra em falta no provérbio.	Quem semeia ventos colhe... tempestades	x	2	2
				vendavais			
				tormentos			
				temporais			
LPP_SP31559	[Q-031]	Quanto mais se tem, mais se quer.	Selecione a palavra em falta no provérbio.	Quanto mais se tem, mais se... quer	x	2	2
				precisa			
				economiza			
				ambiciona			
LPP_JRMC14588	[Q-032]	Gato escaldado, de água fria tem medo.	Selecione a palavra em falta no provérbio.	... escaldado, de água fria tem medo. Gato	x	2	2
				Pato			
				Peixe			
				Felino			
LPP_JRMC04718	[Q-033]	Burro velho não aprende línguas.	Selecione a palavra em falta no provérbio.	Burro velho não aprende ... línguas	x	2	2
				latim	x		
				trava-línguas			
				idioma			
LPP_JRMC23643	[Q-034]	Peixe não puxa carroça.	Selecione a palavra em falta no provérbio.	... não puxa carroça Peixe	x	2	2
				Peixe			
				Burro			
				Cardume			
LPP_SP05805	[Q-035]	As paredes têm ouvidos.	Selecione a palavra em falta no provérbio.	As paredes têm ... ouvidos	x	2	2
				orelha			
				badana			
				tijolo			
LPP_SP28297	[Q-036]	Palavra puxa palavra.	Selecione a palavra em falta no provérbio.	Palavra puxa ... palavra	x	2	2
				argumento			
				frase			
				carroça			
LPP_SP00923	[Q-037]	A experiência é a mãe da ciência.	Selecione a palavra em falta no provérbio.	A ... é a mãe da ciência		2	2

Tabela 3 continuação da página anterior

ID	ID-Q2	Provérbio	Tipo de jogo	Provérbio e resposta	RC	DL-A	DL-MP
				experiência	x		
				sapiência			
				vivência			
PP_AM00166	[Q-038]	A cavalo dado, não se olha ao dente.	Selecione a palavra em falta no provérbio.	A cavalo dado, não se olha ao ... dente		2	2
				pigmento	x		
				pedigree			
				arado			
LPP_JRMC19664	[Q-039]	Apanha-se mais depressa um mentiroso que um coxo.	Selecione a palavra em falta no provérbio.	Apanha-se mais depressa um ... que um coxo		2	2
				mentiroso	x		
				franco			
				falso			
				idoso			
LP_SP02120	[Q-040]	A ocasião faz o ladrão.	Selecione a palavra em falta no provérbio.	A ocasião faz o ... ladrão		2	2
				ladrão	x		
				larvão			
				larápio			
				coração			
LPP_JRMC08097	[Q-041]	Nem tudo o que vem à rede é peixe.	Selecione a palavra em falta no provérbio.	Nem tudo o que vem à ... é peixe rede		2	2
				rede	x		
				linha			
				armadilha			
				baía			
LPP_JRMC03231	[Q-042]	Quem brinca com o fogo queima-se.	Selecione a palavra em falta no provérbio.	Quem brinca com o ... queima-se fogo		2	2
				carvão	x		
				óleo			
				jogo			
LPP_JRMC14130	[Q-043]	Onde há fumo, há fogo.	Selecione a palavra em falta no provérbio.	Onde há ..., há fogo fumo		2	2
				fumo	x		
				nuvem			
				bombeiros			
				fumaceira			
LP_SP07878	[Q-044]	Cantigas, leva-as o vento.	Selecione a palavra em falta no provérbio.	..., leva-as o vento Cantigas		2	2
				Palavras	x		
				Formigas	x		
				Cantões			
LP_SP01307	[Q-045]	A idade não perdoa.	Selecione a palavra em falta no provérbio.	A ... não perdoa idade		2	2
				verdade	x		
				vaidade			
				velhice			

Tabela 3 continuação da página anterior

ID	ID-Q2	Provérbio	Tipo de jogo	Provérbio e resposta	RC	DL-A	DL-MP
LPP_JRMC22577	[Q-046]	A paciência tem limites.	Selecione a palavra em falta no provérbio.	A ... tem limites paciência	x	2	2
				ciência			
				persistência			
				pachorra			
PP_AM02387	[Q-047]	Cão que ladra não morde.	Selecione a palavra em falta no provérbio.	Cão que ladra não ...	x	2	2
				morde			
				foge			
				roe			
				ferre			
LP_SP26828	[Q-048]	O silêncio é a alma do negócio.	Selecione a palavra em falta no provérbio.	O silêncio é a alma do ...		2	2
				negócio	x		
				sossego	x		
				contrato			
				sacerdócio			
LPP_JRMC13673	[Q-049]	A fome é o melhor tempero.	Selecione a palavra em falta no provérbio.	A ... é o melhor tempero		2	2
				fome	x		
				mostarda	x		
				gana			
				salsa			
LPP_JRMC26852	[Q-050]	A saúde não tem preço.	Selecione a palavra em falta no provérbio.	A ... não tem preço		2	2
				saúde	x		
				qualidade			
				cultura			
				lucidez			
LP_SP39774	[Q-051]	Um mal nunca vem só.	Complete o provérbio.	Um mal...		2	2
				nunca vem só	x		
				nunca vem desamparado			
				nunca vem acompanhado			
				vem sempre sozinho			
LPP_JRMC07410	[Q-052]	Diz-me com quem andas, dir-te-ei quem és.	Complete o provérbio.	Diz-me com quem andas, ...		2	2
				dir-te-ei quem és	x		
				dir-te-ei as manhas que tens	x		
				dir-te-ei quem serás			
				e assim serão as demandas			
PP_AM12218	[Q-053]	Todos os caminhos vão dar a Roma.	Complete o provérbio.	Todos os caminhos ...		2	2
				vão dar a Roma	x		
				levam a Roma	x		
				vão dar a Oklahoma			
				vão dar ao rio			
LP_SP35966	[Q-054]	Quem vê caras não vê corações.	Complete o provérbio.	Quem vê caras ...		2	2
				não vê corações	x		
				não vê faras			
				não vê brasões			

Tabela 3 continuação da página anterior

ID	ID-Q2	Provérbio	Tipo de jogo	Provérbio e resposta	RC	DL-A	DL-MP
LP_SP35026	[Q-055]	Quem sabe sabe.	Complete o provérbio.	vê ilusões Quem sabe ... sabe ensina que se gabe não estranha	x	2	2
LPP_JRMC25044	[Q-056]	Há sempre uma primeira vez para tudo.	Complete o provérbio.	Há sempre uma primeira vez ... para tudo para perder para quem fez para o português	x	2	2
LP_SP26802	[Q-057]	O seguro morreu de velho.	Complete o provérbio.	O seguro... morreu de velho morreu centenário morreu adulto morreu de antigo	x	2	2
LP_SP00907	[Q-058]	A esperança é sempre a última a morrer.	Complete o provérbio.	A esperança... é sempre a última a morrer é verde é a última a nascer é sempre a primeira a morrer	x	2	2
LP_SP16461	[Q-059]	Hoje por mim, amanhã por ti.	Complete o provérbio.	Hoje por mim, ... amanhã por ti amanhã por mim amanhã por vós amanhã por mim	x	2	2
PP_AM08381 lpassou2	[Q-060]	O sol quando nasce é para todos.	Complete o provérbio.	hoje sem fim O sol quando nasce ... é para todos é para quem merece não é para todos vem de Este	x	2	2
LPP_JRMC03496	[Q-061]	Quem te avisa, teu amigo é.	Complete o provérbio.	Quem te avisa, ... teu amigo é teu amigo não é não banaliza teu aliado é	x	2	2
GLP_JPM27157	[Q-062]	Um homem prevenido vale por dois.	Complete o provérbio.	Um homem prevenido ... vale por dois jamais será vencido vale por muitos deve ser confido	x	2	2
LP_SP03235	[Q-063]	A vida são dois dias.	Complete o provérbio.	... são dois dias. A vida As festas A existência Esta lida	x	2	2

Tabela 3 continuação da página anterior

ID	ID-Q2	Provérbio	Tipo de jogo	Provérbio e resposta	RC	DL-A	DL-MP
LPP_JRMC09168	[Q-064]	Deitar cedo e cedo erguer, dá saúde e faz crescer.	Complete o provérbio.	... dá saúde e faz crescer		2	2
				Deitar cedo e cedo erguer,	x		
				Deitar tarde e tarde erguer,			
				Dormir bem e bem comer,			
LP_SP26937	[Q-065]	O tempo é dinheiro.	Complete o provérbio.	Levantar antes do amanhecer,		2	2
				O tempo ...			
				é dinheiro	x		
				é fortuna			
				é riqueza			
				é tesouro			
GLP_JPM23744	[Q-066]	Quem sai aos seus não degenera.	Complete o provérbio.	Quem sai aos seus ...		2	2
				não degenera	x		
				não desespera			
				não desonera			
				degenera			
GLP_JPM14466	[Q-067]	Não faças aos outros o que não queres que te façam a ti.	Complete o provérbio.	Não faças aos outros ...		2	2
				o que não queres que te façam a ti	x		
				nada ruim			
				o que não gostas			
				o que não te fazem a ti			
LPP_JRMC04473	[Q-068]	Zangam-se as comadres, descobrem-se as verdades.	Complete o provérbio.	Zangam-se as comadres ...		2	2
				descobrem-se as verdades	x		
				acabam-se as crueldades			
				vão-se as maldades			
				encobrem-se as verdades			
LP_SP33808	[Q-069]	Quem não chora não mama.	Complete o provérbio.	Quem não chora ...		2	2
				não mama	x		
				não ama			
				mija			
				não come			
LP_SP06155	[Q-070]	Azar ao jogo, sorte no amor.	Complete o provérbio.	Azar ao jogo, ...		2	2
				sorte no amor	x		
				azar em tudo			
				é para os principiantes			
LP_SP33850	[Q-071]	Quem não deve não teme.	Complete o provérbio.	Quem não deve ...		2	2
				não teme	x		
				não paga			
				não teve			
				não recebe			

Tabela 3 continuação da página anterior

ID	ID-Q2	Provérbio	Tipo de jogo	Provérbio e resposta	RC	DL-A	DL-MP
LPP_JRMC26104	[Q-072]	Toma lá, dá cá.	Complete o provérbio.	Toma lá ... dá cá	x	2	2
				dá lá			
				leva lá			
				dá acolá			
L_P_SP32267	[Q-073]	Quem com ferro fere com ferro será ferido.	Complete o provérbio.	Quem com ferro fere ... com ferro será ferido não é bom partido	x	2	2
				com ferro não será ferido			
				com ferro será batido			
LPP_JRMC31779	[Q-074]	Quem está vivo sempre aparece.	Complete o provérbio.	Quem está vivo ... sempre aparece	x	2	2
				sempre se mostra			
				é porque não morreu			
				sempre comparece			
LPP_JRMC12091	[Q-075]	Enquanto há vida, há esperança.	Complete o provérbio.	Enquanto há vida ... há esperança	x	2	2
				há lembrança			
				há cobrança			
				há aliança			
LPP_JRMC01585	[Q-076]	Quem tudo quer, tudo perde.	Complete o provérbio.	Quem tudo quer ... tudo perde mais consegue tudo ganha mais perde	x	2	2
				Tal pai ...			
PP_AM12013	[Q-077]	Tal pai, tal filho.	Complete o provérbio.	Tal pai ... tal filho	x	2	2
				tal fruto			
				tal descendente			
				tal neto			
L_P_SP32139	[Q-078]	Quem cala consente.	Complete o provérbio.	Quem cala ... consente	x	2	2
				não desmente			
				permite			
				não mente			
LPP_JRMC27813	[Q-079]	Para quem é, bacalhau basta.	Complete o provérbio.	Para quem é, ... bacalhau basta	x	2	2
				carapau basta			
				nada basta			
				qualquer coisa chega			
PP_AM08501	[Q-080]	Olho por olho, dente por dente.	Complete o provérbio.	Olho por olho ... dente por dente	x	2	2
				mão por mão			
				ficas zarolho			
				gente por gente			

Tabela 3 continuação da página anterior

ID	ID-Q2	Provérbio	Tipo de jogo	Provérbio e resposta	RC	DL-A	DL-MP
LP_SP17338	[Q-081]	Laranja, de manhã é ouro, à tarde prata e à noite mata.	Complete o provérbio.	Laranja, de manhã é ouro, à tarde prata... e à noite mata e à noite é choro e à noite é lata e de noite enfastia	x	2	2
LP_SP14109	[Q-082]	Entre marido e mulher, nunca metas a colher.	Complete o provérbio.	Entre marido e mulher ...		2	2
LPP_JRMC25137	[Q-083]	O fruto proibido é o mais apetecido.	Complete o provérbio.	nunca metas a colher nunca metas o talher nunca se mete o bedelho mete sempre a colher	x		
LP_SP32209	[Q-084]	Quem casa quer casa.	Complete o provérbio.	O fruto proibido ... é o mais apetecido nunca é o mais apetecido é o mais querido é o mais cobiçado Quem casa ... quer casa quer lar quer asa quer teito As más notícias ... chegam depressa chegam sem demora vêm sem pressa vêm devagar	x	2	2
LPP_JRMC21927	[Q-085]	As más notícias chegam depressa.	Complete o provérbio.	Quando mija um português ...		2	1
PP_AM09590	[Q-086]	Quando mija um português, mijam dois ou três.	Complete o provérbio.	mijam dois ou três mija claro e fala direito não mija só uma vez não mija um francês	x		
GLP_JPM03851	[Q-087]	As moscas apanham-se com mel e não com fel.	Complete o provérbio.	As moscas apanham-se ... com mel e não com fel com vinagre com açúcar com fel e não com mel	x	2	1
LP_SP09249	[Q-088]	Com uma mentira se apanha uma verdade.	Complete o provérbio.	Com uma mentira ...		2	2

Tabela 3 continuação da página anterior

ID	ID-Q2	Provérbio	Tipo de jogo	Provérbio e resposta	RC	DL-A	DL-MP
				se apanha uma verdade	x		
				se apanha uma maldade			
				se apanha a falsidade			
				se apanha a realidade			
LP_SP27875	[Q-089]	Os olhos comem mais que a barriga.	Complete o provérbio.	Os olhos comem mais ... que a barriga	x	2	2
				que o ventre			
				que a pança			
				do que os pardais			
LPP_JRMC10014	[Q-090]	Deus dá o frio conforme a roupa.	Complete o provérbio.	Deus dá o frio ... conforme a roupa	x	2	1
				e o Diabo dá o brio			
				conforme o arrepio			
				conforme a vestimenta			
LPP_JRMC25173	[Q-091]	O prometido é devido.	Complete o provérbio.	O prometido ... é devido	x	2	2
				está em dívida			
				é cumprido			
				é garantido			
LPP_JRMC01557	[Q-092]	Quanto mais alto se sobe, maior é a queda.	Complete o provérbio.	Quanto mais alto se sobe ... maior é a queda	x	2	2
				mais depressa se cai			
				menor é a queda			
				maior é o trambolhão			
PP_AM100585	[Q-093]	A palavras loucas, orelhas moucas.	Complete o provérbio.	A palavras loucas ... orelhas moucas	x	2	2
				ouvidos surdos			
				orelhas taralhoulcas			
				ensinanças poucas			
PP_AM10573	[Q-094]	Quem não é por mim, é contra mim.	Complete o provérbio.	Quem não é por mim ... é contra mim	x	2	2
				é porque só sabe de si			
				não é por ninguém			
				é contra ti			
GLP_JPM12101	[Q-095]	Longe da vista, longe do coração.	Complete o provérbio.	Longe da vista, ... longe do coração	x	2	2
				perto da emoção			
				nao há quem resista			
				longe da razão			
LPP_JRMC00270	[Q-096]	Grão a grão enche a galinha o papo.	Complete o provérbio.	... enche a galinha o papo Grão a grão	x	2	2
				Bago a bago	x		
				Comendo pão			
				Com o trapo			
LPP_JRMC11417	[Q-097]	Cada um puxa a brasa à sua sardinha.	Complete o provérbio.	Cada um puxa a brasa ... à sua sardinha	x	2	2

Tabela 3 continuação da página anterior

ID	ID-Q2	Provérbio	Tipo de jogo	Provérbio e resposta	RC	DL-A	DL-MP
				ao seu peixe à sua pinha			
LP_SP21344	[Q-098]	Não há duas sem três.	Complete o provérbio.	para o lado que que convem Não há duas ... sem três para quem não refez sem uma vez sem quatro	x	2	2
LPP_JRM12140	[Q-099]	Quem espera, sempre alcança.	Complete o provérbio.	Quem espera, ... sempre alcança desespera tem de ter paciência colhe	x x	2	2
PP_AM04090	[Q-100]	Em casa de ferreiro, espeto de pau.	Complete o provérbio.	Em casa de ferreiro, ... espeto de pau o espeto é de ferro há sempre um bom faqueiro espeto de madeira	x	2	2

.8 Jogos VITHEA–Kids

Tabela 4: Jogos VITHEA – Kids

Q	ID	Provérbio	Tipo de jogo	Provérbio e resposta
1	LPP_JRMC08894	A curiosidade matou o gato.	Seleciona a palavra que falta no provérbio.	A curiosidade matou o ... gato rato sapo frade
2	LPP_JRMC00270	Grão a grão enche a galinha o papo.	Completa o provérbio.	... enche a galinha o papo Grão a grão De manhã Migalha a migalha Feijão a feijão
2.1		Grão a grão enche a galinha o papo.	Que palavra falta no provérbio?	Grão a grão enche a ... o papo galinha (imagem) menina (imagem) rainha (imagem) joaninha (imagem)
2.2		Grão a grão enche a galinha o papo.	Seleciona a palavra que falta no provérbio.	Grão a grão enche a galinha o ... papo vagão saco alçapão
3	LP_SP01151	A galinha da vizinha é sempre melhor que a minha.	Que palavra falta no provérbio?	A ... da vizinha é sempre melhor que a minha galinha (imagem) sardinha (imagem) casa (imagem)
4	LPP_JRMC03231	Quem brinca com o fogo queima-se.	Seleciona a palavra que falta no provérbio.	Quem brinca com o ... queima-se fogo carvão óleo fole
5	PP_AM00166	A cavalo dado, não se olha ao dente.	Seleciona a palavra que falta no provérbio.	A cavalo dado, não se olha ao ... dente casco dorso estribo
6	PP_AM02387	Cão que ladra não morde.	Seleciona a palavra que falta no provérbio.	Cão que ladra não ... morde foge roe ferra
7	PP_AM04066	Em boca fechada não entra mosca.	Completa o provérbio.	Em boca fechada... não entra mosca as melgas não têm entrada não entra aranha
8	PP_AM03332	De pequenino é que se torce o pepino.	Seleciona a palavra que falta no provérbio.	De pequenino é que se torce o ... pepino nariz menino cebolinho

Tabela 4 continuação da página anterior

Q	ID	Provérbio	Tipo de jogo	Provérbio e resposta
9	LP_SP18450	Mais vale um pássaro na mão do que dois a voar.	Completa o provérbio.	Mais vale um pássaro na mão do que dois ... a voar a pescar na gaiola na árvore
10	LP_SP14109	Entre marido e mulher, nunca metas a colher.	Que palavra falta no provérbio?	Entre marido e mulher, nunca metas a ... colher (imagem) panela (imagem) mão (imagem) faca (imagem)
10.1		Entre marido e mulher, nunca metas a colher.	Completa o provérbio.	Entre marido e mulher, nunca metas ... a colher o nariz o bebelho o bico
11	GLP_JPM23720	Quem ri por último ri melhor	Seleciona a palavra que falta no provérbio.	Quem ri por último ri ... melhor pior bem depois
12	GLP_JPM01343	À noite todos os gatos são pardos.	Seleciona a palavra que falta no provérbio.	À noite todos os ... são pardos gatos ratos patos leopardos
13	GLP_JPM01946	A união faz a força.	Completa o provérbio.	A união faz ... a força a razão a confusão a ação
14	GLP_JPM10955	Gato escaldado de água fria tem medo.	Seleciona a palavra que falta no provérbio.	Gato escaldado de água ... tem medo. fria benta quente fervida
15	LPP_JRMC16719	Quando um burro fala, os outros [a]baixam as orelhas.	Completa o provérbio.	Quando um burro fala, os outros baixam ... as orelhas as cabeças a bola a grimpá
16	GLP_JPM21870	Quem conta um conto acrescenta um ponto.	Seleciona a palavra que falta no provérbio.	Quem conta um conto acrescenta um... ponto confronto capítulo pormenor
17	GLP_JPM22859	Quem não arrisca não petisca.	Seleciona a palavra que falta no provérbio.	Quem não arrisca não ... petisca mordisca lambisca alcanga
18	LP_SP34161	Quem não tem cão caça com gato.	Seleciona a palavra que falta no provérbio.	Quem não tem cão caça com ...

Tabela 4 continuação da página anterior

Q	ID	Provérbio	Tipo de jogo	Provérbio e resposta
				gato
				arco
				falcão
				furão
19	LP_SP22188	Não se fazem omeletes sem ovos	Seleciona a palavra que falta no provérbio.	Não se fazem omeletes sem ... ovos
				sal
				gemas
				galinhas
20	GLP_JPM17109	O fruto proibido é o mais apetecido.	Completa o provérbio.	O fruto proibido ... é o mais apetecido é o mais comido não deve ser comido é o que sabe melhor

.9 Seleção provérbios usuais – Anotador 1

Tabela 5: Seleção provérbios usuais – Anotador 1

ID	Provérbio	Anotador 1
GLP_JPM00323	A cão mordido todos o mordem.	MF
GLP_JPM00394	A César o que é de César e a Deus o que é de Deus.	MF
GLP_JPM00591	A exceção confirma a regra.	MF
GLP_JPM00690	A fome faz sair o lobo do covil.	MF
GLP_JPM00866	A inveja matou Caim.	MF
GLP_JPM01012	A maior vingança é o desprezo.	MF
GLP_JPM01117	A montanha pariu um rato.	MF
GLP_JPM01234	A mulher e a sardinha quer(em)-se (da mais) pequenina.	MF
GLP_JPM01457	A perseverança tudo alcança.	MF
GLP_JPM01936	A última gota é que faz transbordar o copo.	MF
GLP_JPM02053	A vida é cheia de altos e baixos.	MF
GLP_JPM02685	Amor ausente amor para sempre.	MF
GLP_JPM03536	Apressado come cru.	MF
GLP_JPM03733	As águias não caçam moscas.	MF
GLP_JPM03851	As moscas apanham-se com mel e não com fel.	MF
GLP_JPM04987	Cada coisa a seu tempo.	MF
GLP_JPM05049	Cada qual com seu igual.	MF
GLP_JPM05153	Cada um em sua casa é rei.	MF
GLP_JPM05364	Canta Marta depois de farta.	MF
GLP_JPM05928	Chega-te aos bons e serás um deles.	MF
GLP_JPM07402	De nada duvida quem nada sabe.	MF
GLP_JPM07880	Deus ajuda a quem muito madruga.	MF
GLP_JPM08418	Do prato à boca se perde a sopa.	MF
GLP_JPM08650	Duro com duro não faz bom muro.	MF
GLP_JPM09598	Enquanto o pau vai e vem folgam as costas.	MF
GLP_JPM10827	Gaivotas em terra, tempestade no mar.	MF
GLP_JPM11039	Governa-se a boca conforme a bolsa.	MF
GLP_JPM11135	Guarda-te do cão que não ladra e do homem que não fala.	MF
GLP_JPM11223	Há mar e mar, há ir e voltar.	MF
GLP_JPM12035	Língua comprida, mentira maior.	MF
GLP_JPM12101	Longe da vista, longe do coração.	MF
GLP_JPM12483	Mais vale cair em graça (do) que ser engraçado.	MF
GLP_JPM12484	Mais vale calar que mal falar.	MF
GLP_JPM12583	Mais vale perder um minuto na vida que a vida num minuto.	MF
GLP_JPM12625	Mais vale saúde que dinheiro.	MF
GLP_JPM12627	Mais vale sê-lo que parecê-lo.	MF
GLP_JPM12681	Mais vale um pássaro na mão (do) que dois a voar.	MF
GLP_JPM13438	Morra Marta, (mas) morra farta.	MF
GLP_JPM14272	Não cuspas para o ar que te pode cair (o cuspo) na cara.	MF
GLP_JPM14466	Não faças aos outros o que não queres que te façam a ti.	MF
GLP_JPM14580	Não há carne sem osso, nem fruta sem caroço.	MF
GLP_JPM14715	Não há madeira sem nó.	MF
GLP_JPM14831	Não há pastor sem rebanho.	MF
GLP_JPM14896	Não há roca sem o seu fuso.	MF
GLP_JPM15044	Não metas o nariz onde não és (fores) chamado.	MF
GLP_JPM15248	Não se fala em corda em casa de enforcado.	MF
GLP_JPM15823	Nem só de pão vive o homem.	MF
GLP_JPM15897	Nem tudo que luz é ouro.	MF
GLP_JPM15994	Ninguém é indispensável.	MF
GLP_JPM16021	Ninguém foge ao seu destino.	MF
GLP_JPM16065	Ninguém sabe para o que nasce.	MF
GLP_JPM16364	Numa mulher não se bate nem com uma flor.	MF
GLP_JPM16396	Nunca Deus fecha uma porta que não abra outra.	MF
GLP_JPM16406	Nunca é tarde para aprender.	MF
GLP_JPM17025	O dormir é meio sustento.	MF
GLP_JPM17391	O mel não é para a boca do asno.	MF
GLP_JPM17534	O olho do dono engorda o cavalo.	MF
GLP_JPM17762	O que é bom não dura.	MF
GLP_JPM18023	O que tem de ser tem muita força.	MF
GLP_JPM18153	O silêncio é de ouro, a palavra é de prata.	MF
GLP_JPM18177	O Sol quando nasce é para todos.	MF
GLP_JPM18275	O trabalho do menino é pouco mas quem o despreza é louco.	MF
GLP_JPM18479	Onde canta o galo não canta a galinha.	MF
GLP_JPM18552	Onde muitos mandam ninguém obedece.	MF
GLP_JPM18696	Os erros pagam-se caro.	MF
GLP_JPM18887	Ou oito ou oitenta.	MF
GLP_JPM19089	Palavras não enchem barriga.	MF
GLP_JPM19125	Panela que muitos mexem não toma tempero.	MF
GLP_JPM20066	Pobreza não é vergonha.	MF
GLP_JPM20068	Pobreza não é vileza.	MF
GLP_JPM20232	Por morrer uma andorinha não acaba a Primavera.	MF
GLP_JPM21220	Quanto mais prima, mais se lhe arrima.	MF
GLP_JPM21426	Quem alheio veste na praça o despe.	MF
GLP_JPM21684	Quem canta seus males (seu mal) espanta.	MF

Tabela 5 continuação da página anterior

ID	Provérbio	Anotador 1
GLP_JPM21886	Quem cospe para cima na cara lhe cai.	MF
GLP_JPM22379	Quem faz um cesto faz um cento.	MF
GLP_JPM22422	Quem gaba a noiva é o noivo.	MF
GLP_JPM22563	Quem mais jura mais mente.	MF
GLP_JPM22696	Quem mexe no mel sempre se lambe.	MF
GLP_JPM23018	Quem não poupa não tem.	MF
GLP_JPM23263	Quem nasceu para burro nunca chega a cavalo.	MF
GLP_JPM23537	Quem procura ilusão acaba desiludido.	MF
GLP_JPM23720	Quem ri por último ri melhor.	MF
GLP_JPM23744	Quem sai aos seus não degenera.	MF
GLP_JPM23934	Quem semeia ventos colhe tempestades.	MF
GLP_JPM24290	Quem tem vergonha morre de fome.	MF
GLP_JPM24858	Saber esperar é uma (grande) virtude.	MF
GLP_JPM25404	Se queres um bom conselho, pede-o ao velho.	MF
GLP_JPM25500	Segredo melhor guardado é o que a ninguém é revelado.	MF
GLP_JPM26824	Três foi a conta que Deus fez.	MF
GLP_JPM27123	Um é pouco, dois é bom, três é de mais.	MF
GLP_JPM27157	Um homem prevenido vale por dois.	MF
GLP_JPM27243	Uma andorinha só não faz o verão.	MF
GLP_JPM27444	Vai à fava enquanto a ervilha (não) enche.	MF
GLP_JPM27563	Vão-se os anéis mas fiquem-se os dedos.	MF
LP_SP00014	A acha sai à racha.	MF
LP_SP00098	A água silenciosa é a mais perigosa.	MF
LP_SP00244	A beleza está nos olhos de quem a vê.	MF
LP_SP00363	A bom gato, bom rato.	MF
LP_SP00389	A brincar, a brincar, é que o macaco fez um filho à mãe.	MF
LP_SP00538	A carne é fraca.	MF
LP_SP00707	A corda arrebenta sempre do lado mais fraco.	MF
LP_SP00772	A descer todos os santos ajudam.	MF
LP_SP00907	A esperança é sempre a última a morrer.	MF
LP_SP00923	A experiência é a mãe da ciência.	MF
LP_SP00969	A fé move montanhas.	MF
LP_SP01151	A galinha da vizinha é sempre melhor que a minha.	MF
LP_SP01307	A idade não perdoa.	MF
LP_SP01370	A intenção é que conta.	MF
LP_SP01427	A justiça tarda, mas não falta.	MF
LP_SP01435	A lã nunca pesou ao carneiro.	MF
LP_SP01710	A mentira tem pernas curtas.	MF
LP_SP01723	A mil chegarás, de dois mil não passarás.	MF
LP_SP01808	A morte não escolhe idades.	MF
LP_SP02120	A ocasião faz o ladrão.	MF
LP_SP02147	A ovelha busca (procura) a sua parelha.	MF
LP_SP02191	A paixão cega a razão.	MF
LP_SP02369	A preguiça morreu de sede à beira da água (ao pé de um rio).	MF
LP_SP02383	A pressa é inimiga da perfeição.	MF
LP_SP02907	A sorte de uns é o azar de outros.	MF
LP_SP03027	A união faz a força.	MF
LP_SP03157	A verdade sempre vem à tona.	MF
LP_SP03193	A vida é bela, o que é preciso é saber vivê-la.	MF
LP_SP03235	A vida são dois dias.	MF
LP_SP03257	A vingança serve-se fria.	MF
LP_SP03372	Abril, águas mil.	MF
LP_SP03522	Afoga-se mais gente em vinho do que em água.	MF
LP_SP03886	Amanhã é outro dia.	MF
LP_SP03987	Amigo não empata amigo.	MF
LP_SP04012	Amigos, amigos, negócios à parte.	MF
LP_SP04021	Amigos dos meus amigos, meus amigos são.	MF
LP_SP04071	Amor com amor se paga.	MF
LP_SP04186	Amores velhos nunca se esquecem.	MF
LP_SP04335	Ano de abelhas, ano de ovelhas.	MF
LP_SP04413	Antes a lã se perca que a ovelha.	MF
LP_SP04473	Antes dar que receber.	MF
LP_SP04634	Antes que cases olha (vê) o que fazes.	MF
LP_SP04646	Antes que o mal cresça corta-se(-lhe) a cabeça.	MF
LP_SP04728	Antes viver pobre que morrer rico.	MF
LP_SP05175	Aprende até morrer.	MF
LP_SP05508	As árvores morrem de pé.	MF
LP_SP05542	As coisas devem chamar-se pelos nomes.	MF
LP_SP05560	As coisas são eternas enquanto duram.	MF
LP_SP05797	As palavras são como as cerejas, atrás duma vêm as outras.	MF
LP_SP05805	As paredes têm ouvidos.	MF
LP_SP05840	As rosas caem e os espinhos ficam.	MF
LP_SP06066	Atrás da cruz se esconde o diabo.	MF
LP_SP06096	Atrás do tempo, tempo vem.	MF
LP_SP06155	Azar ao jogo, sorte no amor.	MF
LP_SP06390	Beleza não põe mesa.	MF
LP_SP06813	Boda molhada, boda abençoada.	MF
LP_SP07161	Burro morto, cevada ao rabo.	MF
LP_SP07373	Cada macaco no seu galho.	MF
LP_SP07505	Cada um colhe conforme (como) semeia.	MF
LP_SP07605	Cada um sabe as linhas com que se cose.	MF
LP_SP07616	Cada um sabe onde lhe aperta o sapato.	MF
LP_SP07620	Cada um safa-se como pode.	MF

Tabela 5 continuação da página anterior

ID	Provérbio	Anotador 1
LP_SP07878	Cantigas, leva-as o vento.	MF
LP_SP08180	Casa que não é ralhada não é bem governada.	MF
LP_SP08207	Casa varrida e mulher penteada, parece bem e não custa nada.	MF
LP_SP09135	Com o tempo, tudo se cura.	MF
LP_SP09171	Com papas e bolos, se enganam os tolos.	MF
LP_SP09249	Com uma mentira se apanha uma verdade.	MF
LP_SP09527	Companhia de três é má rês.	MF
LP_SP09786	Cré com cré, lé com lé.	MF
LP_SP09802	Cria fama e deita-te na cama.	MF
LP_SP09970	Da discussão nasce a luz.	MF
LP_SP10300	De amor se vive, de amor se morre.	MF
LP_SP10317	De boa árvore, bom fruto.	MF
LP_SP11145	Depressa e bem não há quem.	MF
LP_SP11415	Deus não dorme.	MF
LP_SP12085	Do mais feio botão nasce a mais linda rosa.	MF
LP_SP12385	Dos males o menor.	MF
LP_SP13133	Em boca fechada entra mosca.	MF
LP_SP13970	Enquanto se canta, não se assobia.	MF
LP_SP13987	Enquanto um burro fala, o outro baixa as orelhas.	MF
LP_SP14109	Entre marido e mulher, nunca metas a colher.	MF
LP_SP14181	Errar é humano.	MF
LP_SP14916	Faze o que (eu) te digo e não o que eu faço.	MF
LP_SP15389	Filho de peixe sabe nadar.	MF
LP_SP15425	Filhos casados, cuidados dobrados.	MF
LP_SP15575	Formosura pouco dura.	MF
LP_SP15841	Ganha boa fama e deita-te a dormir.	MF
LP_SP16133	Guarda hoje para amanhã.	MF
LP_SP16278	Há gente para tudo.	MF
LP_SP16357	Há remédio para tudo menos para a morte.	MF
LP_SP16461	Hoje por mim, amanhã por ti.	MF
LP_SP17188	Junta-se o útil ao agradável.	MF
LP_SP17338	Laranja, de manhã é ouro, à (de) tarde (ao meio dia) prata e à noite mata.	MF
LP_SP17632	Lua de mel, lua de fel.	MF
LP_SP17890	Mais caro é o dado que o comprado.	MF
LP_SP18083	Mais vale a prática que a gramática.	MF
LP_SP18320	Mais vale prevenir do que remediar.	MF
LP_SP18387	Mais vale tarde do que nunca.	MF
LP_SP18431	Mais vale um gosto na vida que cem mil réis na algibeira.	MF
LP_SP18453	Mais vale um pé no travão que dois no caixão.	MF
LP_SP18497	Mais vale uma mão inchada que uma enxada na mão.	MF
LP_SP18561	Mal de muitos consolo é.	MF
LP_SP18763	Mão vai, mão vem.	MF
LP_SP18774	Mãos frias, coração quente.	MF
LP_SP19485	Mocidade desprevenida, velhice arrependida.	MF
LP_SP19720	Muita parra, pouca uva.	MF
LP_SP19902	Muito sabe o rato, mas mais sabe o gato.	MF
LP_SP20003	Mulher barbuda, Deus nos acuda.	MF
LP_SP20012	Mulher bem casada não tem sogra nem cunhada.	MF
LP_SP20138	Mulher honesta não tem ouvidos.	MF
LP_SP20141	Mulher honrada deve ser calada.	MF
LP_SP20228	Mulher ri quando pode e chora quando quer.	MF
LP_SP20339	Na (pela) capa se conhece o dono.	MF
LP_SP20709	Não cantes vitória antes de tempo.	MF
LP_SP20752	Não compres gato por lebre.	MF
LP_SP20842	Não deites foguetes, antes da festa.	MF
LP_SP20866	Não des conselho sem que to peçam.	MF
LP_SP20943	Não é a pancada da vara que amadurece a azeitona.	MF
LP_SP21004	Não é o bater das asas que faz a águia.	MF
LP_SP21038	Não é por o galo cantar que há-de madrugar.	MF
LP_SP21167	Não gastes cera com ruins defuntos.	MF
LP_SP21270	Não há cego que se veja nem torto que se conheça.	MF
LP_SP21310	Não há crime perfeito.	MF
LP_SP21344	Não há duas sem três.	MF
LP_SP21380	Não há fome sem fartura.	MF
LP_SP21685	Não há regra sem excepção, nem mulher sem senão.	MF
LP_SP21756	Não há trigo sem joio.	MF
LP_SP21809	Não julgues os cabos pelos começos.	MF
LP_SP21811	Não julgues para não ser julgado.	MF
LP_SP21874	Não meças tudo pela mesma medida.	MF
LP_SP21990	Não ponhas o carro à frente dos bois.	MF
LP_SP22127	Não se caçam lebres tocando tambor.	MF
LP_SP22130	Não se começa a casa pelo telhado.	MF
LP_SP22153	Não se deve cuspir no prato onde se comeu.	MF
LP_SP22158	Não se deve despir um santo para vestir outro.	MF
LP_SP22188	Não se faz(em) omelete(s) sem ovos.	MF
LP_SP22226	Não se morre de amor.	MF
LP_SP22274	Não se pode ter tudo.	MF
LP_SP22419	Não te rias do mal do teu vizinho, que o teu já vem a caminho.	MF
LP_SP22869	Nem o avô morre nem a gente almoça.	MF
LP_SP23048	Nem tudo o que parece é.	MF
LP_SP23050	Nem tudo o que se diz se escreve.	MF
LP_SP23198	Ninguém é de ferro.	MF
LP_SP23199	Ninguém é de ninguém.	MF

Tabela 5 continuação da página anterior

ID	Provérbio	Anotador 1
LP_SP23226	Ninguém empobrece por ter dado muito.	MF
LP_SP23545	No meio é que está a virtude.	MF
LP_SP23826	Nunca digas: Desta água não beberei.	MF
LP_SP24014	O abismo chama o abismo.	MF
LP_SP24368	O bom cão não ladra em falso.	MF
LP_SP24545	O cão e o menino vão onde sentem mimo.	MF
LP_SP24685	O coração nunca se engana.	MF
LP_SP24902	O dinheiro não compra a felicidade.	MF
LP_SP25454	O mais difícil é o primeiro passo.	MF
LP_SP25671	O menino e o borracho, põe-lhe Deus a mão por baixo.	MF
LP_SP25700	O meu é meu, o teu é nosso.	MF
LP_SP25902	O pão do pobre cai sempre com a manteiga para o lado de baixo.	MF
LP_SP26301	O que está feito não tem remédio.	MF
LP_SP26364	O que mal começa mal acaba.	MF
LP_SP26405	O que não é por mim é contra mim.	MF
LP_SP26415	O que não mata engorda.	MF
LP_SP26802	O seguro morreu de velho.	MF
LP_SP26825	O seu a seu dono.	MF
LP_SP26828	O silêncio é a alma do negócio.	MF
LP_SP26937	O tempo é dinheiro.	MF
LP_SP26963	O tempo não perdoa.	MF
LP_SP26991	O tempo vai e não volta.	MF
LP_SP27613	Os amigos são para as ocasiões.	MF
LP_SP27734	Os fins justificam os meios.	MF
LP_SP27875	Os olhos comem mais que a barriga.	MF
LP_SP27890	Os opostos atraem-se.	MF
LP_SP28212	Paga o justo pelo pecador.	MF
LP_SP28297	Palavra puxa palavra.	MF
LP_SP28831	Para quem não sabe ler um pingo é letra.	MF
LP_SP28904	Para vilão, vilão e meio.	MF
LP_SP29071	Pau que nasce torto tarde ou nunca se endireita.	MF
LP_SP29206	Pela boca morre o peixe.	MF
LP_SP29556	Perca-se tudo menos a honra.	MF
LP_SP30017	Por causa de uns, pagam os outros.	MF
LP_SP31368	Quando um cai, todos o pisam.	MF
LP_SP31506	Quanto mais me bates, mais (eu) gosto de ti.	MF
LP_SP31559	Quanto mais se tem, mais se quer.	MF
LP_SP31629	Quatro olhos vêem mais que dois.	MF
LP_SP32139	Quem cala consente.	MF
LP_SP32209	Quem casa quer casa.	MF
LP_SP32254	Quem com cães se deita com pulgas acorda.	MF
LP_SP32267	Quem com ferro fere com ferro será ferido.	MF
LP_SP32395	Quem compra fiado paga dobrado.	MF
LP_SP32663	Quem desdenha quer comprar.	MF
LP_SP33014	Quem está de fora vê (sempre) melhor.	MF
LP_SP33044	Quem fala assim não é gago.	MF
LP_SP33138	Quem feio ama bonito lhe parece.	MF
LP_SP33364	Quem mais alto subir de mais alto vai cair.	MF
LP_SP33370	Quem mais chora menos mija.	MF
LP_SP33397	Quem mais promete menos cumpre.	MF
LP_SP33748	Quem nada não se afoga.	MF
LP_SP33808	Quem não chora não mama.	MF
LP_SP33826	Quem não confia não é de confiar.	MF
LP_SP33850	Quem não deve não teme.	MF
LP_SP34070	Quem não sabe inventa.	MF
LP_SP34192	Quem não tem dinheiro não tem vícios.	MF
LP_SP34246	Quem não tem pé não dá coice.	MF
LP_SP34278	Quem não tem unhas não toca viola.	MF
LP_SP34298	Quem não trabuca não manduca.	MF
LP_SP34461	Quem os meus filhos beija minha boca adoça.	MF
LP_SP34513	Quem parte e reparte fica com a melhor parte.	MF
LP_SP34625	Quem pode pode.	MF
LP_SP34850	Quem quer respeito dá respeito.	MF
LP_SP35026	Quem sabe sabe.	MF
LP_SP35476	Quem tem boca vai a Roma.	MF
LP_SP35566	Quem tem filhos tem cadilhos.	MF
LP_SP35590	Quem tem inimigos não dorme.	MF
LP_SP35600	Quem tem medo compra um cão.	MF
LP_SP35714	Quem tem telhado(s) de vidro não atira pedra(s) ao do vizinho.	MF
LP_SP35811	Quem trabalha de graça é o relógio.	MF
LP_SP35895	Quem vai à guerra dá e leva.	MF
LP_SP35949	Quem vai para o mar avia-se primeiro em terra.	MF
LP_SP35966	Quem vê caras não vê corações.	MF
LP_SP36078	Querer é poder.	MF
LP_SP36394	Ri-se o nu do mal vestido.	MF
LP_SP36423	Roma e Pavia não se fizeram num dia.	MF
LP_SP36698	Santos de casa não fazem milagres.	MF
LP_SP37657	Sem trabalho, nada se consegue.	MF
LP_SP38012	Só se vive uma vez.	MF
LP_SP39012	Todos os sacos têm fundo.	MF
LP_SP39024	Todos querem o que é seu.	MF
LP_SP39664	Um dia da caça, outro do caçador.	MF
LP_SP39774	Um mal nunca vem só.	MF

Tabela 5 continuação da página anterior

ID	Provérbio	Anotador 1
LP_SP39866	Um santo não pode estar em dois altares.	MF
LP_SP40024	Uma imagem vale (mais que) mil palavras.	MF
LP_SP40073	Uma no cravo, outra na ferradura.	MF
LP_SP40454	Vaso ruim não (nunca) quebra.	MF
LP_SP40961	Viver não custa, o que custa é saber viver.	MF
LP_SP41017	Zangas de namorados, amores dobrados.	MF
LPP_JRMC00103	Do uso, nasce o abuso.	MF
LPP_JRMC00155	Quando vires as barbas do vizinho a arder, põe as tuas de molho.	MF
LPP_JRMC00161	A má acção fica com quem a pratica.	MF
LPP_JRMC00193	O que é achado não é roubado.	MF
LPP_JRMC00233	Se Maomé não vai à montanha, a montanha que venha a Maomé.	MF
LPP_JRMC00270	Grão a grão enche a galinha o papo.	MF
LPP_JRMC00375	O corno é o último a saber.	MF
LPP_JRMC00443	A gente só se lembra de Santa Bárbara quando troveja.	MF
LPP_JRMC00450	Na hora do aperto é que se conhecem os amigos.	MF
LPP_JRMC00491	Não se pode agradar a gregos e a troianos.	MF
LPP_JRMC00597	Por medo dos pardais não se deixa de semear cereais.	MF
LPP_JRMC00815	A água lava tudo menos as más línguas.	MF
LPP_JRMC00818	A água o dá, a água o leva.	MF
LPP_JRMC00945	A agulha puxa a linha.	MF
LPP_JRMC01060	Os filhos são a alegria dos pais.	MF
LPP_JRMC01451	Por cada porta que se fecha, outra se abre.	MF
LPP_JRMC01465	Ama gorda, pouco leite.	MF
LPP_JRMC01473	Amanhã também é dia.	MF
LPP_JRMC01557	Quanto mais alto se sobe, maior é a queda.	MF
LPP_JRMC01585	Quem tudo quer, tudo perde.	MF
LPP_JRMC01782	Mais vale um amigo próximo que um parente afastado.	MF
LPP_JRMC02008	Vinho e amigo, o mais antigo.	MF
LPP_JRMC02151	Amor não enche barriga.	MF
LPP_JRMC02255	Não há amor sem dor.	MF
LPP_JRMC02294	O amor e a morte vencem o mais forte.	MF
LPP_JRMC02323	O amor não tem lei.	MF
LPP_JRMC02462	O amor é cego.	MF
LPP_JRMC02533	Não há amor como o primeiro.	MF
LPP_JRMC02660	Os anjos não têm costas.	MF
LPP_JRMC02664	Ano bissexto, ano travesso.	MF
LPP_JRMC02713	Não contes com o ovo no cu da galinha.	MF
LPP_JRMC02719	Não contes os pintos senão depois de nascidos.	MF
LPP_JRMC02722	Não esperes por sapatos de defunto.	MF
LPP_JRMC02763	As aparências iludem.	MF
LPP_JRMC02817	Os homens não se medem aos palmos.	MF
LPP_JRMC02887	Aprende e saberás.	MF
LPP_JRMC02922	Vivendo e aprendendo.	MF
LPP_JRMC03074	Pela obra se conhece o artista.	MF
LPP_JRMC03122	À primeira cavadela, minhoca.	MF
LPP_JRMC03143	Não bebas sem ver, nem assines sem ler.	MF
LPP_JRMC03177	Não deixes caminhos por atalhos.	MF
LPP_JRMC03191	Quem se mete em atalhos, mete-se em trabalhos.	MF
LPP_JRMC03231	Quem brinca com o fogo queima-se.	MF
LPP_JRMC03283	Para a frente é que é o caminho.	MF
LPP_JRMC03345	A pequeno passarinho, pequeno ninho.	MF
LPP_JRMC03496	Quem te avisa, teu amigo é.	MF
LPP_JRMC03679	As bebidas fortes fazem os homens fracos.	MF
LPP_JRMC03781	Antes da sopa molha-se a boca.	MF
LPP_JRMC03819	Se bebes de mais, tropeças e caís.	MF
LPP_JRMC03841	Beleza sem virtude é rosa sem cheiro.	MF
LPP_JRMC03849	Não há bela sem senão.	MF
LPP_JRMC03909	Há males que vêm por bem.	MF
LPP_JRMC03976	Fazer bem sem olhar a quem.	MF
LPP_JRMC04256	O boi em terra alheia é vaca.	MF
LPP_JRMC04470	Quando um não quer, dois não brigam.	MF
LPP_JRMC04473	Zangam-se as comadres, descobrem-se as verdades.	MF
LPP_JRMC04479	A brincar é que as coisas se dizem.	MF
LPP_JRMC04520	Antes asno que me leve que cavalo que me derrube.	MF
LPP_JRMC04718	Burro velho não aprende línguas.	MF
LPP_JRMC04746	Cabelos compridos, ideias curtas.	MF
LPP_JRMC04915	A grande cão, grande osso.	MF
LPP_JRMC05224	Não digas o que sabes sem saber o que dizes.	MF
LPP_JRMC05353	Quem tem calos não se mete em apertos.	MF
LPP_JRMC05525	Quem tem unhas é que toca guitarra.	MF
LPP_JRMC05550	Cara de um, focinho de outro.	MF
LPP_JRMC05715	A casas velhas, portas novas.	MF
LPP_JRMC06259	Virou-se o feitiço contra o feitiço.	MF
LPP_JRMC06296	Cá se fazem, cá se pagam.	MF
LPP_JRMC06402	A cavalo novo, cavaleiro velho.	MF
LPP_JRMC06608	O pior cego é o que não quer ver.	MF
LPP_JRMC06980	O comer e o coçar, o mal é começar.	MF
LPP_JRMC07022	Uma mão lava a outra e ambas o rosto.	MF
LPP_JRMC07204	Onde come um, comem dois.	MF
LPP_JRMC07304	Come para viver, não vivas para comer.	MF
LPP_JRMC07410	Diz-me com quem andas, dir-te-ei quem és.	MF
LPP_JRMC07439	Antes só que mal acompanhado.	MF
LPP_JRMC07565	Quem é de todos não é de ninguém.	MF

Tabela 5 continuação da página anterior

ID	Provérbio	Anotador 1
LPP_JRMC07758	Quem não te conhecer que te compre.	MF
LPP_JRMC07824	Se o conselho fosse bom não se dava, vendia-se.	MF
LPP_JRMC07983	As boas contas fazem os bons amigos.	MF
LPP_JRMC08010	Contas, são contas.	MF
LPP_JRMC08097	Nem tudo o que vem à rede é peixe.	MF
LPP_JRMC08164	A casamento e a baptizado não vás sem ser convidado.	MF
LPP_JRMC08335	Em Roma sê romano.	MF
LPP_JRMC08503	É preciso ver para crer.	MF
LPP_JRMC08696	Quem dorme com criança, acorda molhado.	MF
LPP_JRMC08741	Parir é dor, criar é amor.	MF
LPP_JRMC08748	O crime não compensa.	MF
LPP_JRMC08752	Um crime não justifica outro.	MF
LPP_JRMC08784	Cautela e caldos de galinha nunca fizeram mal a ninguém.	MF
LPP_JRMC08818	A culpa morreu solteira.	MF
LPP_JRMC08894	A curiosidade matou o gato.	MF
LPP_JRMC08951	Cada um dança conforme a música.	MF
LPP_JRMC09168	Deitar cedo e cedo erguer, dá saúde e faz crescer.	MF
LPP_JRMC09240	O que é de mais não presta.	MF
LPP_JRMC09372	Não te fies em cantigas.	MF
LPP_JRMC09515	Mais vale um desengano que viver enganado.	MF
LPP_JRMC09523	Perdido por cem, perdido por mil.	MF
LPP_JRMC09621	Uma desgraça nunca vem só.	MF
LPP_JRMC09652	A porta da rua é a serventia da casa.	MF
LPP_JRMC09675	Não des pérolas a porcos.	MF
LPP_JRMC09710	Quem o seu inimigo poupa, nas mãos lhe morre.	MF
LPP_JRMC09836	Deus dá nozes a quem não tem dentes.	MF
LPP_JRMC09944	Cada um por si e Deus por todos.	MF
LPP_JRMC10014	Deus dá o frio conforme a roupa.	MF
LPP_JRMC10026	A quem Deus não dá filhos, o diabo dá sobrinhos.	MF
LPP_JRMC10182	Não devas a quem deveu, nem sirvas a quem serviu.	MF
LPP_JRMC10236	A falar no diabo e ele a aparecer.	MF
LPP_JRMC10456	Duro de cozer, duro de roer.	MF
LPP_JRMC10491	Não deixes para amanhã o que podes fazer hoje.	MF
LPP_JRMC10848	Não há concórdia sem discussão.	MF
LPP_JRMC10853	Não ensines o padre-nosso ao vigário.	MF
LPP_JRMC10969	Melhor é dobrar que quebrar.	MF
LPP_JRMC10978	O que é doce nunca amargou.	MF
LPP_JRMC11127	Cada maluco com a sua mania.	MF
LPP_JRMC11134	De poeta e de louco, todos têm um pouco.	MF
LPP_JRMC11303	Quem muito dorme pouco aprende.	MF
LPP_JRMC11370	Antes a criança chore que a mãe suspire.	MF
LPP_JRMC11415	Cada um leva a água ao seu moinho.	MF
LPP_JRMC11417	Cada um puxa a brasa à sua sardinha.	MF
LPP_JRMC11489	É pior a emenda que o soneto.	MF
LPP_JRMC11705	Para bom entendedor, meia palavra basta.	MF
LPP_JRMC11751	Errando é que se aprende.	MF
LPP_JRMC11882	Venha o diabo e escolha.	MF
LPP_JRMC11987	Não se apanham trutas a bragas enxutas.	MF
LPP_JRMC12006	Quem semeia, colhe.	MF
LPP_JRMC12013	Gota a gota o mar se esgota.	MF
LPP_JRMC12052	Quando a esmola é grande, o pobre desconfia.	MF
LPP_JRMC12091	Enquanto há vida, há esperança.	MF
LPP_JRMC12126	É preciso dar tempo ao tempo.	MF
LPP_JRMC12137	Quem espera, desespera.	MF
LPP_JRMC12140	Quem espera, sempre alcança.	MF
LPP_JRMC12301	As coisas querem-se com conta, peso e medida.	MF
LPP_JRMC12303	Dois pesos, duas medidas.	MF
LPP_JRMC12313	Entre mortos e feridos alguém há-de escapar.	MF
LPP_JRMC12422	Nem oito nem oitenta.	MF
LPP_JRMC12490	A falar é que a gente se entende.	MF
LPP_JRMC12547	Do dizer ao fazer vai grande distância.	MF
LPP_JRMC12694	Pensa duas vezes antes de falar.	MF
LPP_JRMC12718	Quem diz o que quer, ouve o que não quer.	MF
LPP_JRMC12835	Fama sem proveito faz mal ao peito.	MF
LPP_JRMC12877	Santo da terra não faz milagre.	MF
LPP_JRMC12915	Quem não faz pó, não faz farinha.	MF
LPP_JRMC13033	Dizer é fácil, o difícil é fazer.	MF
LPP_JRMC13673	A fome é o melhor tempero.	MF
LPP_JRMC13747	Para a fome não há pão duro.	MF
LPP_JRMC14015	Cada um tem o seu calcanhar de Aquiles.	MF
LPP_JRMC14018	Dos fracos não reza a História.	MF
LPP_JRMC14130	Onde há fumo, há fogo.	MF
LPP_JRMC14281	A galinha velha faz bom caldo.	MF
LPP_JRMC14347	Galinha que canta quer galo.	MF
LPP_JRMC14416	Na falta de capão, cebola e pão.	MF
LPP_JRMC14418	O bom galo não engorda.	MF
LPP_JRMC14588	Gato escaldado, de água fria tem medo.	MF
LPP_JRMC14656	Gato miador não é bom caçador.	MF
LPP_JRMC14771	Os gostos não se discutem.	MF
LPP_JRMC14873	Quanto maior é a nau, maior a tormenta.	MF
LPP_JRMC14975	Em tempo de guerra não se limpam armas.	MF
LPP_JRMC15804	Ao ruim, ruim e meio.	MF

Tabela 5 continuação da página anterior

ID	Provérbio	Anotador 1
LPP_JRMC15909	Mais vale solteiro andar que mal casar.	MF
LPP_JRMC16168	Numa hora cai a casa.	MF
LPP_JRMC16169	Numa hora não se ganhou Samora.	MF
LPP_JRMC16172	Todos têm a sua hora.	MF
LPP_JRMC16270	Quanto mais uma pessoa se abaixa, mais se lhe vê o cu.	MF
LPP_JRMC16377	Quem não sabe é como quem não vê.	MF
LPP_JRMC16500	Não se pode dar o que não se tem.	MF
LPP_JRMC16509	Ninguém nasce ensinado.	MF
LPP_JRMC16588	Não deixes o certo pelo duvidoso.	MF
LPP_JRMC16683	Cada um é como cada qual.	MF
LPP_JRMC16744	Pelo dedo se conhece o gigante.	MF
LPP_JRMC16770	De mal agradecidos está o Inferno cheio.	MF
LPP_JRMC16864	Se não podes vencer os teus inimigos, junta-te a eles.	MF
LPP_JRMC16990	De boas intenções está o Inferno cheio.	MF
LPP_JRMC17099	Mais vale ser invejado que lastimado.	MF
LPP_JRMC17192	Mais vale jeito que força.	MF
LPP_JRMC17279	Carta na mesa é carta jogada.	MF
LPP_JRMC17484	Quando a cabeça não tem juízo, o corpo é que paga.	MF
LPP_JRMC17612	Mocidade ociosa, velhice trabalhosa.	MF
LPP_JRMC17666	Ladrão que rouba a ladrão, tem cem anos de perdão.	MF
LPP_JRMC17884	Não há crime sem lei.	MF
LPP_JRMC18029	Mais vale andar esfarrapado que algemado.	MF
LPP_JRMC18105	Lisonjas ouvir, orelhas abrir.	MF
LPP_JRMC18569	Não é por muito madrugar que amanhece mais cedo.	MF
LPP_JRMC18601	Mãe, há só uma.	MF
LPP_JRMC18644	Com o mal dos outros posso eu bem.	MF
LPP_JRMC18684	Não há mal que sempre dure, nem bem que se não acabe.	MF
LPP_JRMC18704	O bem soa, o mal voa.	MF
LPP_JRMC18709	O mal alheio dá conselho.	MF
LPP_JRMC18787	Do mal, o menos.	MF
LPP_JRMC19046	Muito pode o galo no seu poleiro.	MF
LPP_JRMC19296	Para grandes males, grandes remédios.	MF
LPP_JRMC19316	Quem não morre da doença, morre da cura.	MF
LPP_JRMC19436	Quem tem cu, tem medo.	MF
LPP_JRMC19495	Não há mel sem fel.	MF
LPP_JRMC19626	Quem conta um conto, acrescenta-lhe um ponto.	MF
LPP_JRMC19664	Apanha-se mais depressa um mentiroso que um coxo.	MF
LPP_JRMC19695	A merda é a mesma, as moscas é que mudam.	MF
LPP_JRMC19704	Cada um tem aquilo que merece.	MF
LPP_JRMC19706	Em terra de cegos quem tem um olho é rei.	MF
LPP_JRMC19724	Quanto tens, quanto vales.	MF
LPP_JRMC20099	Quem pouco tem, pouco lhe basta.	MF
LPP_JRMC20214	Mais vale morte que má sorte.	MF
LPP_JRMC20407	Morte desejada, vida dobrada.	MF
LPP_JRMC20526	Quem está mal que se mude.	MF
LPP_JRMC20616	A mulher e a mula, o pau as cura.	MF
LPP_JRMC20756	Atrás de um grande homem há sempre uma grande mulher.	MF
LPP_JRMC21084	O homem na praça e a mulher em casa.	MF
LPP_JRMC21296	Gordura é formosura, magreza é beleza.	MF
LPP_JRMC21306	Mulher formosa, mulher perigosa.	MF
LPP_JRMC21321	Mulher honrada não tem ouvidos.	MF
LPP_JRMC21451	O mundo acaba para quem morre.	MF
LPP_JRMC21927	As más notícias chegam depressa.	MF
LPP_JRMC21980	Albarde-se o burro à vontade do dono.	MF
LPP_JRMC21982	Aprende a obedecer, aprenderás a mandar.	MF
LPP_JRMC22025	Obra começada, obra acabada.	MF
LPP_JRMC22112	A ociosidade é a mãe de todos os vícios.	MF
LPP_JRMC22192	Cada um no seu ofício.	MF
LPP_JRMC22269	Os olhos são o espelho da alma.	MF
LPP_JRMC22300	A ferro quente, malhar de repente.	MF
LPP_JRMC22319	Há mais marés que marinheiros.	MF
LPP_JRMC22352	Aponta para o alto e acertarás no alvo.	MF
LPP_JRMC22516	Ovelha que berra, bocado que perde.	MF
LPP_JRMC22560	Para ovos frigar temos de os partir.	MF
LPP_JRMC22577	A paciência tem limites.	MF
LPP_JRMC22724	Quem paga adiantado é mal servido.	MF
LPP_JRMC22730	Paga o que deves e saberás o que tens.	MF
LPP_JRMC22757	A pai guardador, filho gastador.	MF
LPP_JRMC23163	Parar é morrer.	MF
LPP_JRMC23168	Antes ser que parecer.	MF
LPP_JRMC23180	Cada panela com seu testo.	MF
LPP_JRMC23275	Águas passadas não movem moinhos.	MF
LPP_JRMC23297	O que lá vai, lá vai.	MF
LPP_JRMC23643	Peixe não puxa carroça.	MF
LPP_JRMC23680	Roupa branca em Janeiro é sinal de pouco dinheiro.	MF
LPP_JRMC23708	A pensar morreu um burro.	MF
LPP_JRMC23863	Para pergunta apressada, resposta pausada.	MF
LPP_JRMC24044	Se a pílula bem soubera, não se dourara por fora.	MF
LPP_JRMC24307	Em casa onde não há pão, todos ralham e ninguém tem razão.	MF
LPP_JRMC24335	Manda quem pode, obedece quem quer.	MF
LPP_JRMC24582	No poupar é que está o ganho.	MF
LPP_JRMC24715	Depois da casa roubada, trancas à porta.	MF
LPP_JRMC25044	Há sempre uma primeira vez para tudo.	MF

Tabela 5 continuação da página anterior

ID	Provérbio	Anotador 1
LPP_JRMC25117	Quem procura, sempre encontra.	MF
LPP_JRMC25137	O fruto proibido é o mais apetecido.	MF
LPP_JRMC25173	O prometido é devido.	MF
LPP_JRMC25338	Para cada ocasião tenha um provérbio sempre à mão.	MF
LPP_JRMC25368	Mais vale prudência que ciência.	MF
LPP_JRMC25556	Muito ralha Maria em casa vazia.	MF
LPP_JRMC25640	Enquanto dormem os gatos, folgam os ratos.	MF
LPP_JRMC25776	Recordar é viver.	MF
LPP_JRMC25884	O rei vai nu.	MF
LPP_JRMC25948	Tudo está bem quando acaba bem.	MF
LPP_JRMC25998	A primeira todos caem, à segunda cai quem quer e à terceira caem os tolos.	MF
LPP_JRMC26003	Quem parte velho, paga novo.	MF
LPP_JRMC26030	Os dados estão lançados.	MF
LPP_JRMC26042	Quem não se dá ao respeito, não é respeitado.	MF
LPP_JRMC26045	Respeita se queres ser respeitado.	MF
LPP_JRMC26104	Toma lá, dá cá.	MF
LPP_JRMC26279	Quem tem a faca e o queijo, corta onde quer.	MF
LPP_JRMC26339	A quem é rico não faltam parentes.	MF
LPP_JRMC26378	Rir é o melhor remédio.	MF
LPP_JRMC26407	Não cabem dois galos no mesmo poleiro.	MF
LPP_JRMC26442	Não há rosa sem espinhos.	MF
LPP_JRMC26602	O saber não ocupa lugar.	MF
LPP_JRMC26852	A saúde não tem preço.	MF
LPP_JRMC26887	Onde não entra o sol, entra o médico.	MF
LPP_JRMC26972	O segredo é a alma do negócio.	MF
LPP_JRMC27070	Nunca digas nunca, nem sempre.	MF
LPP_JRMC27121	As acções ficam com quem as pratica.	MF
LPP_JRMC27135	A carapuça é para quem a enfia.	MF
LPP_JRMC27184	Não sirvas a quem serviu, nem peças a quem pediu.	MF
LPP_JRMC27280	Cada um é para o que nasce.	MF
LPP_JRMC27547	A sombra da noqueira não te deites a dormir.	MF
LPP_JRMC27558	Não se endireita a sombra de uma vara torta.	MF
LPP_JRMC27719	Todos os filhos da puta têm sorte.	MF
LPP_JRMC27731	A boca da barra se perde o navio.	MF
LPP_JRMC27813	Para quem é, bacalhau basta.	MF
LPP_JRMC28191	Há tempo para velar e tempo para descansar.	MF
LPP_JRMC28402	O tempo voa.	MF
LPP_JRMC28583	Quando os porcos bailam, adivinham chuva.	MF
LPP_JRMC28762	Março ventoso, Abril chuvoso.	MF
LPP_JRMC28954	Março marçagão, manhãs de Inverno, tardes de Verão.	MF
LPP_JRMC29919	Dia de São Martinho, castanhas e vinho.	MF
LPP_JRMC30111	Depois da tempestade vem a bonança.	MF
LPP_JRMC30713	O trabalho não mata ninguém.	MF
LPP_JRMC30870	Tristezas não pagam dívidas.	MF
LPP_JRMC30894	Os últimos são os primeiros.	MF
LPP_JRMC31390	Com a verdade me enganas.	MF
LPP_JRMC31511	O hábito não faz o monge.	MF
LPP_JRMC31779	Quem está vivo sempre aparece.	MF
LPP_JRMC32181	O bom vinho faz bom sangue.	MF
PP_AM00166	A cavalo dado, não se olha ao dente.	MF
PP_AM00368	A ladrão de casa, não há chave.	MF
PP_AM00497	À mulher a roca e ao marido a espada.	MF
PP_AM00559	A noite é boa conselheira.	MF
PP_AM00561	À noite todos os gatos são pardos.	MF
PP_AM00585	A palavras loucas, orelhas moucas.	MF
PP_AM00604	A pêra, quando madura, há-de cair.	MF
PP_AM00773	A rico não devas, a pobre não prometas.	MF
PP_AM00782	A roda da fortuna tanto anda como desanda.	MF
PP_AM00785	A roupa suja lava-se em casa.	MF
PP_AM00849	A uns morrem as vacas, a outros parem os bois.	MF
PP_AM00908	Abre um olho para vender e dois para comprar.	MF
PP_AM00992	Água mole em pedra dura, tanto bate até que fura.	MF
PP_AM01186	Amor que pica, sempre fica.	MF
PP_AM01280	Antes burro vivo que letrado morto.	MF
PP_AM01351	Antes pouco que nada.	MF
PP_AM01406	Ao amigo que não é certo, com um olho fechado, com outro aberto.	MF
PP_AM01425	Ao diabo e à mulher, nunca falta que fazer.	MF
PP_AM01544	Aos parvos aparecem os santos.	MF
PP_AM01786	Até ao lavar dos cestos é vindima.	MF
PP_AM01860	Basta uma ovelha ranhosa para perder um rebanho.	MF
PP_AM02201	Cada cabeça, cada sentença.	MF
PP_AM02356	Candeia que vai adiante, alumia duas vezes.	MF
PP_AM02387	Cão que ladra não morde.	MF
PP_AM02691	Com coisas sérias não se brinca.	MF
PP_AM02994	Cresce e aparece.	MF
PP_AM03236	De Espanha nem bom vento, nem bom casamento.	MF
PP_AM03332	De pequenino é que se torce o pepino.	MF
PP_AM03569	Deus escreve direito por linhas tortas.	MF
PP_AM03619	Devagar se vai ao longe.	MF
PP_AM03751	Do bom vinho, bom vinagre.	MF
PP_AM03992	É pelo estômago que se governam os homens.	MF
PP_AM04090	Em casa de ferreiro, espeto de pau.	MF

Tabela 5 continuação da página anterior

ID	Provérbio	Anotador 1
PP_AM04650	Fia-te na Virgem e não corras e verás o tombo que levas.	MF
PP_AM05361	Livros cerrados não fazem letrados.	MF
PP_AM05646	Mais vale puta na cama que na fama.	MF
PP_AM05671	Mais vale ser mulher de ninguém que amante de alguém.	MF
PP_AM05710	Mais vale uma palavra antes que duas depois.	MF
PP_AM05714	Mais vale verde no meu papo que maduro no alheio.	MF
PP_AM05733	Mal com mal se cura.	MF
PP_AM05880	Mau capelão, mau sacristão.	MF
PP_AM05993	Mente quem dá com a língua no dente.	MF
PP_AM06107	Mordedura de cão, cura-se com o pêlo do mesmo cão.	MF
PP_AM06185	Muito atura quem precisa.	MF
PP_AM06224	Muito riso, pouco siso.	MF
PP_AM06404	Na casa onde a mulher manda, até o galo canta fino.	MF
PP_AM06479	Não busques o pão no focinho do cão.	MF
PP_AM06697	Não deixes a ovelha a guardar o lobo.	MF
PP_AM06740	Não é com vinagre que se apanham moscas.	MF
PP_AM06754	Não é o bom bocado para a boca do asno.	MF
PP_AM06889	Não há gosto sem desgosto.	MF
PP_AM06976	Não há pior surdo que o que não quer ouvir.	MF
PP_AM07194	Não sujes a água que hás-de beber.	MF
PP_AM07280	Nasce o trigo conforme o semeiam.	MF
PP_AM07400	Nem tanto ao mar, nem tanto à terra.	MF
PP_AM07459	Ninguém é bom juiz em causa própria.	MF
PP_AM07495	Ninguém se levanta sem primeiro ter caído.	MF
PP_AM07562	No melhor pano cai a nódoa.	MF
PP_AM07600	No tempo quente, refresca o ventre.	MF
PP_AM07648	Nunca faças nada sem consultar a almofada.	MF
PP_AM07734	O bacalhau quer alho.	MF
PP_AM07736	O barato sai caro.	MF
PP_AM07761	O boi luzidio nunca tem fastio.	MF
PP_AM07777	O bom filho à casa torna.	MF
PP_AM07824	O cão com raiva seu dono morde.	MF
PP_AM07839	O casamento e a mortalha, no céu se talha.	MF
PP_AM07883	O diabo não é tão feio como o pintam.	MF
PP_AM07948	O futuro a Deus pertence.	MF
PP_AM08057	O melhor espelho é amigo velho.	MF
PP_AM08171	O primeiro milho é dos pardais.	MF
PP_AM08194	O que arde cura.	MF
PP_AM08244	O que mais custa melhor sabe.	MF
PP_AM08264	O que não tem remédio, remediado está.	MF
PP_AM08501	Olho por olho, dente por dente.	MF
PP_AM08506	Olhos que não vêem, coração que não sente.	MF
PP_AM08615	Os cães ladram e a caravana passa.	MF
PP_AM08768	Palavras leva-as o vento.	MF
PP_AM08824	Pão pão, queijo queijo.	MF
PP_AM08851	Para baixo todos os santos ajudam.	MF
PP_AM08892	Para morrer basta estar vivo.	MF
PP_AM08986	Patrão fora, dia santo na loja.	MF
PP_AM09590	Quando mija um português, mijam dois ou três.	MF
PP_AM09633	Quando o mar bate na rocha quem se lixa é o mexilhão.	MF
PP_AM09711	Quanto mais depressa mais devagar.	MF
PP_AM09715	Quanto mais há mais se gasta.	MF
PP_AM09826	Quem anda à chuva molha-se.	MF
PP_AM09913	Quem boa cama fizer nela se deitará.	MF
PP_AM10024	Quem comer a carne, roa os ossos.	MF
PP_AM10048	Quem corre por gosto, não cansa.	MF
PP_AM10059	Quem dá aos pobres, empresta a Deus.	MF
PP_AM10069	Quem dá e torna a tirar, ao inferno vai parar.	MF
PP_AM10077	Quem dá o pão, dá a educação.	MF
PP_AM10140	Quem diz a verdade, não merece castigo.	MF
PP_AM10293	Quem foi ao ar, perdeu o lugar.	MF
PP_AM10485	Quem muito fala, pouco acerta.	MF
PP_AM10516	Quem muito escolhe, pouco acerta.	MF
PP_AM10547	Quem não arrisca, não petisca.	MF
PP_AM10573	Quem não é por mim, é contra mim.	MF
PP_AM10575	Quem não é visto, não é lembrado.	MF
PP_AM10588	Quem não gosta, não come.	MF
PP_AM10632	Quem não quer ser lobo, não lhe vista a pele.	MF
PP_AM10690	Quem não tem cão, caça com um gato.	MF
PP_AM10719	Quem não trabalha, não come.	MF
PP_AM10816	Quem pergunta, quer saber.	MF
PP_AM11177	Quem tem fome, cardos come.	MF
PP_AM11222	Quem tem um, não tem nenhum.	MF
PP_AM11302	Quem vier atrás, que feche a porta.	MF
PP_AM11377	Rei morto, rei posto.	MF
PP_AM11522	São mais as vozes que as nozes.	MF
PP_AM11935	Sogra e madrasta, só o nome basta.	MF
PP_AM12013	Tal pai, tal filho.	MF
PP_AM12039	Tantas vezes vai o cântaro à fonte que lá deixa a asa.	MF
PP_AM12080	Tão ladrão é o que entra como o que fica à porta.	MF
PP_AM12218	Todos os caminhos vão dar a Roma.	MF
PP_AM12377	Um burro carregado de livros é um doutor.	MF
PP_AM12420	Um olho no prato, outro no gato.	MF

Tabela 5 continuação da página anterior

ID	Provérbio	Anotador 1
PP_AM12619	Velhos são os trapos.	MF
PP_AM12763	Voz do povo, voz de Deus.	MF
PP_AM12764	Vozes de burro não chegam aos céus.	MF

.10 Seleção provérbios usuais – Anotador 2

Tabela 6: Seleção provérbios usuais – Anotador 2

ID	Provérbio	Anotador 2
GLP_JPM01117	A montanha pariu um rato.	mf
GLP_JPM01234	A mulher e a sardinha quer(em)-se (da mais) pequenina.	mf
GLP_JPM02053	A vida é cheia de altos e baixos.	mf
GLP_JPM02685	Amor ausente amor para sempre.	mf
GLP_JPM03851	As moscas apanham-se com mel e não com fel.	mf
GLP_JPM05059	Cada qual é como Deus o fez.	mf
GLP_JPM05063	Cada qual é para o que nasce.	mf
GLP_JPM06995	Da mulher e da sardinha a mais pequenina.	mf
GLP_JPM08739	E dito e feito.	mf
GLP_JPM09598	Enquanto o pau vai e vem folgam as costas.	mf
GLP_JPM10827	Gaivotas em terra, tempestade no mar.	mf
GLP_JPM11085	Grão a grão enche a galinha o papo.	mf
GLP_JPM11681	Já a formiga tem catarro.	mf
GLP_JPM12101	Longe da vista, longe do coração.	mf
GLP_JPM12681	Mais vale um pássaro na mão (do) que dois a voar.	mf
GLP_JPM14466	Não faças aos outros o que não queres que te façam a ti.	mf
GLP_JPM15897	Nem tudo que luz é ouro.	mf
GLP_JPM18023	O que tem de ser tem muita força.	mf
GLP_JPM18131	O seguro morreu de velho.	mf
GLP_JPM18153	O silêncio é de ouro, a palavra é de prata.	mf
GLP_JPM18887	Ou oito ou oitenta.	mf
GLP_JPM21240	Quanto mais se vive mais se aprende.	mf
GLP_JPM21450	Quem anda à chuva molha-se.	mf
GLP_JPM21684	Quem canta seus males (seu mal) espanta.	mf
GLP_JPM22379	Quem faz um cesto faz um cento.	mf
GLP_JPM22563	Quem mais jura mais mente.	mf
GLP_JPM22859	Quem não arrisca não petisca.	mf
GLP_JPM23720	Quem ri por último ri melhor.	mf
GLP_JPM23744	Quem sai aos seus não degenera.	mf
GLP_JPM23867	Quem se não sente não é filho de boa gente.	mf
GLP_JPM23934	Quem semeia ventos colhe tempestades.	mf
GLP_JPM24373	Quem tudo quer tudo perde.	mf
GLP_JPM25523	Seguro morreu de velho.	mf
GLP_JPM26824	Três foi a conta que Deus fez.	mf
GLP_JPM27157	Um homem prevenido vale por dois.	mf
LP_SP00040	A afeição cega a razão.	mf
LP_SP00244	A beleza está nos olhos de quem a vê.	mf
LP_SP00538	A carne é fraca.	mf
LP_SP00772	A descer todos os santos ajudam.	mf
LP_SP00907	A esperança é sempre a última a morrer.	mf
LP_SP00923	A experiência é a mãe da ciência.	mf
LP_SP00969	A fé move montanhas.	mf
LP_SP01151	A galinha da vizinha é sempre melhor que a minha.	mf
LP_SP01307	A idade não perdoa.	mf
LP_SP01370	A intenção é que conta.	mf
LP_SP01427	A justiça tarda, mas não falta.	mf
LP_SP01808	A morte não escolhe idades.	mf
LP_SP02120	A ocasião faz o ladrão.	mf
LP_SP02191	A paixão cega a razão.	mf
LP_SP02369	A preguiça morreu de sede à beira da água (ao pé de um rio).	mf
LP_SP02383	A pressa é inimiga da perfeição.	mf
LP_SP03027	A união faz a força.	mf
LP_SP03235	A vida são dois dias.	mf
LP_SP03257	A vingança serve-se fria.	mf
LP_SP03372	Abril, águas mil.	mf
LP_SP03509	Afeição cega a razão.	mf
LP_SP03886	Amanhã é outro dia.	mf
LP_SP03987	Amigo não empata amigo.	mf
LP_SP04012	Amigos, amigos, negócios à parte.	mf
LP_SP04071	Amor com amor se paga.	mf
LP_SP04523	Antes dobrar que quebrar.	mf
LP_SP05508	As árvores morrem de pé.	mf
LP_SP05805	As paredes têm ouvidos.	mf
LP_SP06155	Azar ao jogo, sorte no amor.	mf
LP_SP06390	Beleza não põe mesa.	mf
LP_SP07040	Boniteza não se põe na mesa.	mf
LP_SP07373	Cada macaco no seu galho.	mf
LP_SP07425	Cada qual é como Deus o fez.	mf
LP_SP07605	Cada um sabe as linhas com que se cose.	mf
LP_SP07610	Cada um sabe das linhas com que se cose.	mf
LP_SP07878	Cantigas, leva-as o vento.	mf
LP_SP07931	Cão que não ladra morde.	mf
LP_SP08107	Casa de ferreiro, espeto de pau.	mf
LP_SP08185	Casa roubada, trancas à porta.	mf
LP_SP08394	Cavalo dado não se olha o dente.	mf
LP_SP09135	Com o tempo, tudo se cura.	mf
LP_SP09171	Com papas e bolos, se enganam os tolos.	mf
LP_SP09249	Com uma mentira se apanha uma verdade.	mf

Tabela 6 continuação da página anterior

ID	Provérbio	Anotador 2
LP_SP09786	Cré com cré, lé com lé.	mf
LP_SP10661	De pensar morreu um burro.	mf
LP_SP10915	Deitar cedo e cedo erguer dá saúde e faz crescer.	mf
LP_SP11145	Depressa e bem não há quem.	mf
LP_SP11335	Deus dá nozes a quem não tem dentes.	mf
LP_SP11415	Deus não dorme.	mf
LP_SP13970	Enquanto se canta, não se assobia.	mf
LP_SP14109	Entre marido e mulher, nunca metas a colher.	mf
LP_SP14181	Errar é humano.	mf
LP_SP14384	Esta vida são dois dias.	mf
LP_SP14973	Fazer bem e não olhar a quem.	mf
LP_SP15389	Filho de peixe sabe nadar.	mf
LP_SP15431	Filhos criados, trabalhos dobrados.	mf
LP_SP16278	Há gente para tudo.	mf
LP_SP16357	Há remédio para tudo menos para a morte.	mf
LP_SP16461	Hoje por mim, amanhã por ti.	mf
LP_SP17338	Laranja, de manhã é ouro, à (de) tarde (ao meio dia) prata e à noite mata.	mf
LP_SP17632	Lua de mel, lua de fel.	mf
LP_SP18264	Mais vale o jeito que a força.	mf
LP_SP18320	Mais vale prevenir do que remediar.	mf
LP_SP18387	Mais vale tarde do que nunca.	mf
LP_SP18774	Mãos frias, coração quente.	mf
LP_SP19388	Mija um português, mijam dois ou três.	mf
LP_SP19720	Muita parra, pouca uva.	mf
LP_SP21344	Não há duas sem três.	mf
LP_SP21685	Não há regra sem exceção, nem mulher sem senão.	mf
LP_SP23271	Ninguém nasce ensinado.	mf
LP_SP24014	O abismo chama o abismo.	mf
LP_SP24552	O cão que ladra não morde.	mf
LP_SP24902	O dinheiro não compra a felicidade.	mf
LP_SP25700	O meu é meu, o teu é nosso.	mf
LP_SP26354	O que lá vai lá vai.	mf
LP_SP26405	O que não é por mim é contra mim.	mf
LP_SP26502	O que os olhos não vêem, o coração não sente.	mf
LP_SP26802	O seguro morreu de velho.	mf
LP_SP26828	O silêncio é a alma do negócio.	mf
LP_SP26933	O tempo e a maré não esperam por ninguém.	mf
LP_SP26937	O tempo é dinheiro.	mf
LP_SP27571	Onde vai o carro, vão os bois.	mf
LP_SP27613	Os amigos são para as ocasiões.	mf
LP_SP27734	Os fins justificam os meios.	mf
LP_SP27875	Os olhos comem mais que a barriga.	mf
LP_SP28273	Paixão cega a razão.	mf
LP_SP28297	Palavra puxa palavra.	mf
LP_SP29206	Pela boca morre o peixe.	mf
LP_SP30017	Por causa de uns, pagam os outros.	mf
LP_SP31559	Quanto mais se tem, mais se quer.	mf
LP_SP31629	Quatro olhos vêem mais que dois.	mf
LP_SP32113	Quem cá as faz cá as paga.	mf
LP_SP32139	Quem cala consente.	mf
LP_SP32209	Quem casa quer casa.	mf
LP_SP32267	Quem com ferro fere com ferro será ferido.	mf
LP_SP32663	Quem desdenha quer comprar.	mf
LP_SP33044	Quem fala assim não é gago.	mf
LP_SP33138	Quem feio ama bonito lhe parece.	mf
LP_SP33364	Quem mais alto subir de mais alto vai cair.	mf
LP_SP33370	Quem mais chora menos mija.	mf
LP_SP33384	Quem mais fala mais erra.	mf
LP_SP33808	Quem não chora não mama.	mf
LP_SP33850	Quem não deve não teme.	mf
LP_SP34123	Quem não se sente não é filho de boa gente.	mf
LP_SP34139	Quem não te conhecer que te compre.	mf
LP_SP34192	Quem não tem dinheiro não tem vícios.	mf
LP_SP34513	Quem parte e reparte fica com a melhor parte.	mf
LP_SP34625	Quem pode pode.	mf
LP_SP35026	Quem sabe sabe.	mf
LP_SP35408	Quem te avisa, teu amigo é.	mf
LP_SP35476	Quem tem boca vai a Roma.	mf
LP_SP35949	Quem vai para o mar avia-se primeiro em terra.	mf
LP_SP35966	Quem vê caras não vê corações.	mf
LP_SP36078	Querer é poder.	mf
LP_SP36423	Roma e Pavia não se fizeram num dia.	mf
LP_SP36698	Santos de casa não fazem milagres.	mf
LP_SP38012	Só se vive uma vez.	mf
LP_SP39490	Tudo o que vier é ganho.	mf
LP_SP39774	Um mal nunca vem só.	mf
LP_SP40592	Venha o diabo e escolha.	mf
LP_SP41017	Zangas de namorados, amores dobrados.	mf
LPP_JRMC00161	A má acção fica com quem a pratica.	mf
LPP_JRMC00255	A grão e grão enche a galinha o papo.	mf
LPP_JRMC00270	Grão a grão enche a galinha o papo.	mf
LPP_JRMC00649	Até o lavar dos cestos é vindima.	mf
LPP_JRMC01557	Quanto mais alto se sobe, maior é a queda.	mf

Tabela 6 continuação da página anterior

ID	Provérbio	Anotador 2
LPP_JRMC01559	Quanto mais se tem, mais se quer.	mf
LPP_JRMC01585	Quem tudo quer, tudo perde.	mf
LPP_JRMC02151	Amor não enche barriga.	mf
LPP_JRMC02462	O amor é cego.	mf
LPP_JRMC02660	Os anjos não têm costas.	mf
LPP_JRMC02763	As aparências iludem.	mf
LPP_JRMC02817	Os homens não se medem aos palmos.	mf
LPP_JRMC02922	Vivendo e aprendendo.	mf
LPP_JRMC03122	A primeira cavadela, minhoca.	mf
LPP_JRMC03231	Quem brinca com o fogo queima-se.	mf
LPP_JRMC03283	Para a frente é que é o caminho.	mf
LPP_JRMC03492	Quem avisa, amigo é.	mf
LPP_JRMC03496	Quem te avisa, teu amigo é.	mf
LPP_JRMC03849	Não há bela sem senão.	mf
LPP_JRMC03909	Há males que vêm por bem.	mf
LPP_JRMC04473	Zangam-se as comadres, descobrem-se as verdades.	mf
LPP_JRMC04718	Burro velho não aprende línguas.	mf
LPP_JRMC05669	O caro é barato e o barato sai caro.	mf
LPP_JRMC06245	Quando não se tem cabeça, o corpo é que paga.	mf
LPP_JRMC06400	A cavalo dado não se olha o dente.	mf
LPP_JRMC06608	O pior cego é o que não quer ver.	mf
LPP_JRMC06846	Em Abril águas mil coadas por um funil.	mf
LPP_JRMC07022	Uma mão lava a outra e ambas o rosto.	mf
LPP_JRMC07410	Diz-me com quem andas, dir-te-ei quem és.	mf
LPP_JRMC07439	Antes só que mal acompanhado.	mf
LPP_JRMC07757	Quem não o conhecer que o compre.	mf
LPP_JRMC07983	As boas contas fazem os bons amigos.	mf
LPP_JRMC07992	Boas contas fazem os bons amigos.	mf
LPP_JRMC08010	Contas, são contas.	mf
LPP_JRMC08097	Nem tudo o que vem à rede é peixe.	mf
LPP_JRMC08164	A casamento e a baptizado não vás sem ser convidado.	mf
LPP_JRMC08748	O crime não compensa.	mf
LPP_JRMC08784	Cautela e caldos de galinha nunca fizeram mal a ninguém.	mf
LPP_JRMC08818	A culpa morreu solteira.	mf
LPP_JRMC08894	A curiosidade matou o gato.	mf
LPP_JRMC08951	Cada um dança conforme a música.	mf
LPP_JRMC09168	Deitar cedo e cedo erguer, dá saúde e faz crescer.	mf
LPP_JRMC09621	Uma desgraça nunca vem só.	mf
LPP_JRMC09652	A porta da rua é a serventia da casa.	mf
LPP_JRMC09762	Quem nasce torto, tarde ou nunca se endireita.	mf
LPP_JRMC09836	Deus dá nozes a quem não tem dentes.	mf
LPP_JRMC09944	Cada um por si e Deus por todos.	mf
LPP_JRMC10014	Deus dá o frio conforme a roupa.	mf
LPP_JRMC10236	A falar no diabo e ele a aparecer.	mf
LPP_JRMC10489	Meu dito, meu feito.	mf
LPP_JRMC10491	Não deixes para amanhã o que podes fazer hoje.	mf
LPP_JRMC10647	O meu é meu, o teu é nosso.	mf
LPP_JRMC10969	Melhor é dobrar que quebrar.	mf
LPP_JRMC11127	Cada maluco com a sua mania.	mf
LPP_JRMC11134	De poeta e de louco, todos têm um pouco.	mf
LPP_JRMC11303	Quem muito dorme pouco aprende.	mf
LPP_JRMC11370	Antes a criança chore que a mãe suspire.	mf
LPP_JRMC11417	Cada um puxa a brasa à sua sardinha.	mf
LPP_JRMC11489	É pior a emenda que o soneto.	mf
LPP_JRMC11705	Para bom entendedor, meia palavra basta.	mf
LPP_JRMC11882	Venha o diabo e escolha.	mf
LPP_JRMC11987	Não se apanham trutas a bragas enxutas.	mf
LPP_JRMC12006	Quem semeia, colhe.	mf
LPP_JRMC12091	Enquanto há vida, há esperança.	mf
LPP_JRMC12137	Quem espera, desespera.	mf
LPP_JRMC12140	Quem espera, sempre alcança.	mf
LPP_JRMC12313	Entre mortos e feridos alguém há-de escapar.	mf
LPP_JRMC12422	Nem oito nem oitenta.	mf
LPP_JRMC12490	A falar é que a gente se entende.	mf
LPP_JRMC12547	Do dizer ao fazer vai grande distância.	mf
LPP_JRMC12659	O comer e o falar, o mal é começar.	mf
LPP_JRMC12693	Pelo falar é que a gente se entende.	mf
LPP_JRMC13033	Dizer é fácil, o difícil é fazer.	mf
LPP_JRMC13673	A fome é o melhor tempero.	mf
LPP_JRMC14130	Onde há fumo, há fogo.	mf
LPP_JRMC14588	Gato escaldado, de água fria tem medo.	mf
LPP_JRMC14743	Gostos não se discutem.	mf
LPP_JRMC14771	Os gostos não se discutem.	mf
LPP_JRMC14873	Quanto maior é a nau, maior a tormenta.	mf
LPP_JRMC15804	Ao ruim, ruim e meio.	mf
LPP_JRMC15900	Um homem prevenido vale por dois.	mf
LPP_JRMC16377	Quem não sabe é como quem não vê.	mf
LPP_JRMC16683	Cada um é como cada qual.	mf
LPP_JRMC16770	De mal agradecidos está o Inferno cheio.	mf
LPP_JRMC17279	Carta na mesa é carta jogada.	mf
LPP_JRMC17484	Quando a cabeça não tem juízo, o corpo é que paga.	mf
LPP_JRMC17666	Ladrão que rouba a ladrão, tem cem anos de perdão.	mf

Tabela 6 continuação da página anterior

ID	Provérbio	Anotador 2
LPP_JRMC18564	A quem madruga Deus ajuda.	mf
LPP_JRMC18601	Mãe, há só uma.	mf
LPP_JRMC18684	Não há mal que sempre dure, nem bem que se não acabe.	mf
LPP_JRMC18787	Do mal, o menos.	mf
LPP_JRMC19296	Para grandes males, grandes remédios.	mf
LPP_JRMC19436	Quem tem cu, tem medo.	mf
LPP_JRMC19664	Apanha-se mais depressa um mentiroso que um coxo.	mf
LPP_JRMC19695	A merda é a mesma, as moscas é que mudam.	mf
LPP_JRMC19706	Em terra de cegos quem tem um olho é rei.	mf
LPP_JRMC20526	Quem está mal que se mude.	mf
LPP_JRMC21084	O homem na praça e a mulher em casa.	mf
LPP_JRMC21296	Gordura é formosura, magreza é beleza.	mf
LPP_JRMC21505	Antes pouco que nada.	mf
LPP_JRMC21853	À noite todos os gatos são pardos.	mf
LPP_JRMC21927	As más notícias chegam depressa.	mf
LPP_JRMC21980	Albarde-se o burro à vontade do dono.	mf
LPP_JRMC22112	A ociosidade é a mãe de todos os vícios.	mf
LPP_JRMC22190	Cada qual com seu ofício.	mf
LPP_JRMC22246	Mais vêem quatro olhos que dois.	mf
LPP_JRMC22577	A paciência tem limites.	mf
LPP_JRMC23163	Parar é morrer.	mf
LPP_JRMC23297	O que lá vai, lá vai.	mf
LPP_JRMC23643	Peixe não puxa carroça.	mf
LPP_JRMC23708	A pensar morreu um burro.	mf
LPP_JRMC24335	Manda quem pode, obedece quem quer.	mf
LPP_JRMC24582	No poupar é que está o ganho.	mf
LPP_JRMC24715	Depois da casa roubada, trancas à porta.	mf
LPP_JRMC25044	Há sempre uma primeira vez para tudo.	mf
LPP_JRMC25117	Quem procura, sempre encontra.	mf
LPP_JRMC25136	Nada sabe tanto como o fruto proibido.	mf
LPP_JRMC25137	O fruto proibido é o mais apetecido.	mf
LPP_JRMC25173	O prometido é devido.	mf
LPP_JRMC25640	Enquanto dormem os gatos, folgam os ratos.	mf
LPP_JRMC25884	O rei vai nu.	mf
LPP_JRMC26003	Quem parte velho, paga novo.	mf
LPP_JRMC26030	Os dados estão lançados.	mf
LPP_JRMC26104	Toma lá, dá cá.	mf
LPP_JRMC26442	Não há rosa sem espinhos.	mf
LPP_JRMC26506	A roupa suja lava-se em casa.	mf
LPP_JRMC26602	O saber não ocupa lugar.	mf
LPP_JRMC26852	A saúde não tem preço.	mf
LPP_JRMC27121	As acções ficam com quem as pratica.	mf
LPP_JRMC27261	Enquanto se canta não se assobia.	mf
LPP_JRMC27277	Cada qual é para o que nasce.	mf
LPP_JRMC27280	Cada um é para o que nasce.	mf
LPP_JRMC27335	Mais vale que sobeje do que falte.	mf
LPP_JRMC27719	Todos os filhos da puta têm sorte.	mf
LPP_JRMC27813	Para quem é, bacalhau basta.	mf
LPP_JRMC28267	O tempo vai e não volta.	mf
LPP_JRMC28390	O tempo é dinheiro.	mf
LPP_JRMC28398	O tempo tudo cura.	mf
LPP_JRMC28402	O tempo voa.	mf
LPP_JRMC28991	Abril, águas mil.	mf
LPP_JRMC29870	Dia de São Martinho, mata o teu porco e prova o teu vinho.	mf
LPP_JRMC29872	Dia de São Martinho, vai à adega e prova o teu vinho.	mf
LPP_JRMC29919	Dia de São Martinho, castanhas e vinho.	mf
LPP_JRMC29921	Dia de São Martinho, lume, castanhas e vinho.	mf
LPP_JRMC30111	Depois da tempestade vem a bonança.	mf
LPP_JRMC31511	O hábito não faz o monge.	mf
LPP_JRMC31779	Quem está vivo sempre aparece.	mf
LPP_JRMC32004	Quem as faz que as pague.	mf
PP_AM00088	A bom entendedor, meia palavra basta.	mf
PP_AM00166	A cavalo dado, não se olha ao dente.	mf
PP_AM00559	A noite é boa conselheira.	mf
PP_AM00561	À noite todos os gatos são pardos.	mf
PP_AM00585	A palavras loucas, orelhas moucas.	mf
PP_AM00601	A pensar morreu um burro.	mf
PP_AM00785	A roupa suja lava-se em casa.	mf
PP_AM01351	Antes pouco que nada.	mf
PP_AM01399	Ao bom entendedor meia palavra basta.	mf
PP_AM01786	Até ao lavar dos cestos é vindima.	mf
PP_AM01828	Bago a bago, enche a galinha o papo.	mf
PP_AM02356	Candeia que vai adiante, alumia duas vezes.	mf
PP_AM02387	Cão que ladra não morde.	mf
PP_AM02700	Com fogo não se brinca.	mf
PP_AM02994	Cresce e aparece.	mf
PP_AM03541	Deus ajuda a quem muito madruga.	mf
PP_AM03569	Deus escreve direito por linhas tortas.	mf
PP_AM03618	Devagar que tenho pressa.	mf
PP_AM03619	Devagar se vai ao longe.	mf
PP_AM04090	Em casa de ferreiro, espeto de pau.	mf
PP_AM04117	Em casa onde não há pão, todos ralham e ninguém tem razão.	mf

Tabela 6 continuação da página anterior

ID	Provérbio	Anotador 2
PP_AM04258	Em terra de cegos, quem tem um olho é rei.	mf
PP_AM04650	Fia-te na Virgem e não corras e verás o tombo que levas.	mf
PP_AM05020	Hoje por mim amanhã por ti.	mf
PP_AM05383	Lua a de Janeiro e amor o primeiro.	mf
PP_AM06110	Morra o gato, morra farto.	mf
PP_AM06531	Não degenera quem sai aos seus.	mf
PP_AM06740	Não é com vinagre que se apanham moscas.	mf
PP_AM06917	Não há mal que sempre dure, nem bem que se não acabe.	mf
PP_AM06976	Não há pior surdo que o que não quer ouvir.	mf
PP_AM07158	Não se é bom juiz em causa própria.	mf
PP_AM07400	Nem tanto ao mar, nem tanto à terra.	mf
PP_AM07734	O bacalhau quer alho.	mf
PP_AM07736	O barato sai caro.	mf
PP_AM07777	O bom filho à casa torna.	mf
PP_AM07959	O hábito não faz o monge.	mf
PP_AM08171	O primeiro milho é dos pardais.	mf
PP_AM08194	O que arde cura.	mf
PP_AM08264	O que não tem remédio, remediado está.	mf
PP_AM08381	O sol quando nasce é para todos.	mf
PP_AM08501	Olho por olho, dente por dente.	mf
PP_AM08506	Olhos que não vêem, coração que não sente.	mf
PP_AM08615	Os cães ladram e a caravana passa.	mf
PP_AM08824	Pão pão, queijo queijo.	mf
PP_AM08851	Para baixo todos os santos ajudam.	mf
PP_AM08892	Para morrer basta estar vivo.	mf
PP_AM09506	Quando a esmola é grande, o santo desconfia.	mf
PP_AM09590	Quando mija um português, mijam dois ou três.	mf
PP_AM09682	Quando uma porta se fecha outra se abre.	mf
PP_AM09711	Quanto mais depressa mais devagar.	mf
PP_AM09712	Quanto mais fala, mais erra.	mf
PP_AM09715	Quanto mais há mais se gasta.	mf
PP_AM09826	Quem anda à chuva molha-se.	mf
PP_AM09913	Quem boa cama fizer nela se deitará.	mf
PP_AM10048	Quem corre por gosto, não cansa.	mf
PP_AM10069	Quem dá e torna a tirar, ao inferno vai parar.	mf
PP_AM10140	Quem diz a verdade, não merece castigo.	mf
PP_AM10293	Quem foi ao ar, perdeu o lugar.	mf
PP_AM10516	Quem muito escolhe, pouco acerta.	mf
PP_AM10547	Quem não arrisca, não petisca.	mf
PP_AM10573	Quem não é por mim, é contra mim.	mf
PP_AM10751	Quem o feio ama, bonito lhe parece.	mf
PP_AM11177	Quem tem fome, cardos come.	mf
PP_AM11302	Quem vier atrás, que feche a porta.	mf
PP_AM11377	Rei morto, rei posto.	mf
PP_AM11522	São mais as vozes que as nozes.	mf
PP_AM11919	Sobre melão, vinho de tostão.	mf
PP_AM12013	Tal pai, tal filho.	mf
PP_AM12039	Tantas vezes vai o cântaro à fonte que lá deixa a asa.	mf
PP_AM12218	Todos os caminhos vão dar a Roma.	mf
PP_AM12278	Trás de mim virá, quem bom me fará.	mf
PP_AM12377	Um burro carregado de livros é um doutor.	mf
PP_AM12420	Um olho no prato, outro no gato.	mf
PP_AM12619	Velhos são os trapos.	mf
PP_AM12763	Voz do povo, voz de Deus.	mf