

# ***Corpus oral de PL2: um novo recurso para a investigação e ensino***<sup>1</sup>

*Isabel Almeida Santos, Isabel Pereira, Cristina Martins, Ana Cristina Macário  
Lopes, Conceição Carapinha, Antonino Silva*  
Faculdade de Letras da Universidade de Coimbra / CELGA-ILTEC

## **Abstract:**

The relevance of empirical data for sustaining research on second language learning is undeniable. Given the strong reliance of the quality of research results on the quality of the data on which these are based, the careful organization of learners' corpora is paramount. The PL2 Oral Corpus project aims to provide researchers with oral production data elicited according to the protocol described, thus allowing for descriptions of learners' interlanguages, across different areas of Portuguese grammar and usage, as well as a sustained definition of PL2 speakers' proficiency levels. This data base may also ground and validate didactic and pedagogical projects.

**Keywords:** Portuguese as a foreign language (PFL), learners' corpora, interlanguage, PFL teaching and learning

**Palavras-chave:** português língua não materna, corpora de produções de aprendentes, interlíngua, aprendizagem e ensino de PL2

## **1. Introdução**

### **1.1. Corpora: potencialidades**

Enquanto acervos de dados autênticos, os corpora fornecem uma base empírica para os trabalhos de descrição de estruturas e de usos linguísticos e respetiva análise, permitindo simultaneamente fundamentar discussões de natureza teórica (Nascimento, 2002: 601). Embora as suas primeiras aplicações tenham, inicialmente, ocorrido na área da lexicografia, são muitos os domínios em cujo estudo se consideram, na atualidade, as potencialidades dos recursos deste tipo.

---

<sup>1</sup> A execução deste trabalho contou com financiamento da FCT, no âmbito do [PEst-OE/LIN/UI0287/2014](http://dx.doi.org/10.21747/2183-9077/rapla32).



A convicção é a de que uma Linguística de Corpus (cujo domínio de aplicação se afigura, então, no limite, tão diverso quanto as necessidades dos seus utilizadores) pode fornecer, a diferentes áreas, «a better *means* of doing things» (O’Keeffe & McCarthy, 2010: 7).

No escopo dos domínios abrangidos encontram-se quer as diversas vertentes de funcionamento da língua (descrição de estruturas), quer as problemáticas da sua aquisição e processamento, bem como as especificidades dos seus usos social e contextualmente determinados.

A aquisição / aprendizagem e o desenvolvimento de competências em língua não materna (L<sup>2</sup>) são, precisamente, áreas em cujo estudo a análise de *corpora* se afigura fundamental. Na verdade, um *corpus* de produções de aprendentes de L2 que integre os dados de caracterização do respetivo perfil pode ser utilizado para sustentar empiricamente a investigação em torno das características da interlíngua, tal como ela se apresenta em determinados e distintos níveis de proficiência; avaliar a importância do conhecimento da língua materna ou de outras línguas não maternas; analisar a forma e extensão com que são visíveis fenómenos de transferência (McNery & Xiao, 2011: 372); dar conta da importância da instrução formal no processo. Por outro lado, há que considerar a aplicabilidade dos resultados da análise de *corpora*, por exemplo, no domínio do ensino, da tradução, da comunicação intercultural e das tecnologias da fala (O’Keeffe & McCarthy, 2010: 7).

## 1.2. O Projeto

O “*Corpus* Oral de PL2” é um projeto em desenvolvimento no âmbito do Centro de Estudos de Linguística Geral e Aplicada (CELGA-ILTEC), da Faculdade de Letras da Universidade de Coimbra. O seu objetivo é a constituição de uma base de dados de produções orais de aprendentes adultos de PL2 inseridos em contexto instrucional em Portugal e a disponibilizar livremente sob a forma de ficheiros áudio<sup>3</sup>.

---

<sup>2</sup> Por razões relacionadas com os objetivos aqui assumidos, prescindimos, neste trabalho, da distinção terminológica e da discussão que a esse propósito se pode desenvolver, entre *língua estrangeira* e *língua segunda*; Assim, a sigla L2 é usada para referir, genericamente, a relação dos aprendentes com uma língua não materna.

<sup>3</sup> As entrevistas já efetuadas apresentam uma duração média de 28:35 minutos; para a captação de voz recorreu-se a gravadores *Olympus WS-811* e *Marantz Professional PMD660*, equipados com microfone interno.



Note-se que dados de PL2 estão já disponíveis, em Portugal, num conjunto de diferentes *corpora*: “*Corpus* de Aquisição de L2 - CAL2”, coordenado por Ana Madeira; “*Corpus* de Produções Escritas de Aprendentes de PL2 - PEAPL2”, coordenado por Cristina Martins (cf. Martins, 2013); “Dados de Aprendizagem de Português Língua Estrangeira”, com coordenação de Isabel Leiria. Trata-se, no entanto, em todos os casos referidos, de *corpora* escritos, resultantes da integração de produções escritas ou da transcrição de produções orais.

Não obstante as suas valências, de que são prova os diversos trabalhos de investigação que a eles têm recorrido, tais materiais apresentam, por esse facto, interesse muito limitado tanto para o estudo das áreas segmental e prosódica como para a investigação no domínio pragmático. Ao conjunto de dados já existente, o “*Corpus* Oral de Português L2” acrescenta, então, um acervo, inovador, de produções **orais**<sup>4</sup>.

O Protocolo que orienta essa recolha (e que descreveremos, com detalhe, neste trabalho) foi concebido de modo a obter os materiais adequados para a descrição (modular ou sistemática) da(s) interlíngua(s) dos aprendentes não só no domínio fónico (Archibald, 1988), mas também nos domínios gramatical, pragmático-discursivo e lexical. Por seu turno, a estrutura da amostra, diversificada em função do nível de proficiência (QECRL, 2001) e da língua materna (LM) do aprendente, viabiliza quer a análise dos efeitos de transferência, quer, ainda, o estudo das relações entre o processo de aprendizagem formal e o desenvolvimento das competências em PL2. Assim, há ainda que considerar o valor dos dados no domínio do ensino de português a aprendentes não nativos e, concretamente, no apoio à conceção e elaboração de materiais instrucionais (como gramáticas pedagógicas e manuais escolares) adequados a diferentes perfis de aprendentes. Este projeto permitirá, pois, obter também dados indispensáveis ao desenvolvimento de um projeto mais abrangente, “Para uma Gramática do PLE”, em que se pretende elaborar materiais instrucionais e *guidelines* para docentes, com base em descrições fundamentadas de interlínguas de aprendentes.

---

4 Um projeto com algumas afinidades (“*Corpus* de Português Língua Estrangeira/Língua Segunda”) desenvolve-se, neste momento, no Centro de Estudos de Linguística da Universidade de Lisboa (cf. <http://www.chuLuL.pt/pt/Investigacao/547>). Afigura-se, portanto, num curto prazo, uma alteração substancial no que diz respeito à natureza dos dados empíricos disponíveis para a investigação no domínio de PL2.



## 2. Construção do *corpus*

### 2.1. Sujeitos

Os dados que constituem a base a disponibilizar são recolhidos junto de estudantes de PL2 que se encontrem a frequentar os cursos ou unidades curriculares que a Faculdade de Letras da Universidade de Coimbra oferece nesse âmbito: concretamente, estes informantes frequentam o Curso Anual ou de Férias de *Língua e Cultura Portuguesa para Estrangeiros* ou as unidades curriculares de *Língua Portuguesa para Erasmus*<sup>5</sup>. Trata-se, portanto, de aprendentes adultos de PL2 que combinam a aprendizagem em contexto formal e experiências de imersão / inserção.

Na seleção dos informantes consideram-se, como já antecipámos, o nível de aprendizagem definido pelo curso ou unidade curricular frequentado e as línguas maternas dos indivíduos (LM). Recolhem-se, assim, dados linguísticos de aprendentes que se localizam entre o nível A1 e o nível C1 (+) e que se repartem por uma diversidade considerável de LM tipologicamente diferenciadas e com origens geográficas dispersas (até ao momento já participaram aprendentes de origem europeia, americana, asiática e africana).

Toda a informação biográfica, sociolinguística e curricular relevante é fornecida, por escrito, pelo informante, que preenche uma ficha de dados individuais e assina uma declaração de consentimento informado. Usamos, neste projeto, o modelo concebido e utilizado no âmbito do “*Corpus* de Produções Escritas de Aprendentes de PL2 - PEAPL2” (cf. <http://www.uc.pt/fluc/rcpl2/metodologia/> e <http://www.uc.pt/fluc/rcpl2/dados/>). No entanto, dada a existência de características fónicas diferenciadoras no universo de língua portuguesa, acrescentou-se, à matriz de base utilizada, um pequeno conjunto de questões explicitamente destinadas a obter dados sobre o contacto de cada aprendente com as diferentes modalidades nacionais do português.

### 2.2. Protocolo

De modo a assegurar a comparabilidade dos dados, o mesmo questionário foi apresentado a todos os sujeitos participantes na recolha, independentemente do nível frequentado por cada um e

---

<sup>5</sup> Note-se que estas últimas são concebidas para receber estudantes de Programas de Mobilidade de todas as Faculdades da Universidade de Coimbra. Para informações de pormenor sobre a história e perfil de cada um destes cursos ou unidades curriculares, cf. Pereira, Martins e Santos (no prelo) ou Rio-Torto (2014: 30-38).



da respetiva LM. No entanto, a dificuldade que, na fase exploratória do projeto, os informantes do nível A1 manifestaram em duas das tarefas apresentadas levou a uma ligeira adaptação desses exercícios, tendo sido criadas versões mais curtas (com exclusão das estruturas ou sequências mais complexas) para este segmento de aprendentes.

O questionário propriamente dito organiza-se em 3 secções fundamentais, com formatos distintos e a partir das quais se obtêm dados com valências também distintas. É, então, proposto ao aprendente um conjunto de seis exercícios, que inclui quer tarefas de produção, quer tarefas de leitura. Como já adiantámos e evidenciaremos a seguir, assim se assegura o interesse dos dados não só para a sustentação de análises nos planos segmental, prosódico e ortoépico, mas também para a fundamentação empírica de estudos de tipo morfossintático, lexical e semântico, bem como para a investigação nos domínios das competências sociolinguística e pragmático-discursiva (QECRL, 2001).

As diferentes tarefas, todas elas de realização oral, organizam-se, então, ao longo daquilo que podemos designar como um *continuum* de formalidade / artificialidade, que progride de um grau menor para um grau maior de atenção e concentração na situação comunicativa (cf. Figura 1).

Secção	Natureza dos dados	Tarefa
1	Produção oral	1 - <i>Entrevista semiestruturada</i>
2		2 - <i>Elicitação de atos ilocutórios</i>
		3 - <i>Construção de um texto narrativo a partir de uma sequência de imagens</i>
		4 - <i>Nomeação de figuras a partir de um suporte pictórico</i>
3	Leitura oral	5 - <i>Leitura de texto</i>
		6.1 / 6.2 - <i>Leitura de listas de palavras</i>

Figura 1: Estrutura do questionário



### 2.2.1. Questionário: Secção 1

A Secção 1 é preenchida com a realização de uma *Entrevista semiestruturada*. Nesse momento do contacto, o aprendente, em interação com o investigador, é instado a relatar experiências, a descrever sensações e emoções e a emitir opiniões; a relação do informante com a língua e a cultura portuguesas constitui o tema motivador. Ao investigador, intencionalmente não interventivo, compete apresentar estímulos, evitando sempre interromper o discurso do informante.

Obtêm-se, por esta via, materiais que, relevantes para a análise da competência comunicativa do aprendente numa perspetiva global, assumem particular interesse para trabalhos sobre o desenvolvimento das competências lexical e semântica, mas também morfológica e sintática em PL2. A duração estimada para esta parte da abordagem é de, aproximadamente, 5 minutos.

### 2.2.2. Questionário: Secção 2

Na Secção 2 volta a motivar-se a produção de discurso, mas, agora, propondo-se ao informante três tarefas distintas. Na Tarefa 2 (*Elicitação de atos ilocutórios*), o aprendente é instado a simular situações que induzem a expressão de: i) um pedido, ii) um convite ou sugestão; iii) uma censura; iv) um agradecimento; v) um pedido de desculpas; vi) um elogio ou reconhecimento. Em todos os casos, se propõe ao informante que formule o ato ilocutório considerando diferentes interlocutores, com os quais estabelecerá relações de proximidade ou de afastamento: por exemplo, pergunta-se ao aprendente como é que formularia o pedido de uma garrafa de água num café ao empregado que aí trabalha ou de que modo pediria, a um irmão, que lhe passasse o sumo durante a refeição familiar.

Aos aprendentes de nível A1 não se apresentam as situações relativas à expressão da censura e do elogio ou reconhecimento, por não ser previsível o domínio das competências sociolinguísticas, pragmático-discursivas e funcionais necessárias para este efeito no nível de aprendizagem em questão (cf. QECRL, 2001: 169-184). Esta é, então, uma tarefa do questionário que se apresenta sob duas configurações ligeiramente diferentes, em termos de extensão e complexidade.



A inclusão de um exercício com este formato resulta da constatação de que a competência pragmática é mais difícil de adquirir do que a competência linguística e que os desvios pragmáticos podem ser mais sérios e causar mais danos à relação interpessoal do que os de outra natureza (Thomas, 1983). Observa-se, ainda, que são frequentes os casos em que alunos de nível mais avançado exibem uma competência pragmática ainda não acurada (Barron, 2003).

A pesquisa no âmbito da Pragmática da Interlíngua tem vindo, por isso, a acentuar a investigação acerca dos atos ilocutórios; assim, a eliciação de dados, como os descritos, permitir-nos-á, por um lado, obter informações relativamente às áreas mais críticas do processo da sua aprendizagem; por outro, e considerando que se trata de dois domínios sensíveis – o dos atos diretivos e o dos atos expressivos –, os dados permitirão a abordagem de tópicos como os relacionados com padrões de uso relativos a contextos simétricos e assimétricos, fórmulas de cortesia, grau de convencionalização dos atos.

Já na Tarefa 3, ao aprendente é solicitada a *Construção de um texto narrativo* a partir da visualização de uma sequência de imagens, que deverá apreciar, num primeiro momento, em silêncio. Ao informante é dito que nenhuma informação pode ser fornecida pela remissão, por gestos, para o suporte pictórico e que é necessário verbalizar todos os pormenores da sequência narrativa. Como no caso anterior, também este exercício foi apresentado sob duas formas, que diferem no que diz respeito ao número de imagens apresentadas e, portanto, no grau de complexidade da estrutura narrativa a construir; naturalmente, aos alunos do nível A1 é apresentada a versão mais curta.

Além da sua relevância para a análise das seleções lexicais e das construções sintáticas ativadas pelos aprendentes, os dados obtidos mediante esta solicitação permitirão aferir os desvios mais significativos na construção da coesão textual, nomeadamente no que diz respeito à boa construção de cadeias de referência através da anáfora pronominal, bem como à sequencialização adequada de tempos verbais.

As duas últimas tarefas descritas possibilitarão, em suma, aferir as “competências comunicativas em língua” dos informantes e, mais concretamente, as suas competências pragmáticas (QERCL, 2001: 156). No âmbito destas últimas, a eliciação de dados permitirá avaliar os conhecimentos do utilizador/aprendente relativamente à competência funcional, isto é,



à capacidade de formular enunciados adequados a fins muito específicos, e à competência discursiva, ou seja, à capacidade de organizar sequencialmente um conjunto de frases de modo coeso e coerente (QECRL, 2001: 174-178).

Finalmente, na última Tarefa desta segunda secção do questionário propõe-se um exercício onomasiológico, concretamente, de *Nomeação de figuras*, a partir de um suporte pictórico. Ao contrário dos exercícios anteriores, pretende-se, agora, obter, um conjunto pré-definido de vocábulos isolados (18 nomes e adjetivos), cuja seleção, como explicitaremos em 2.3., decorre essencialmente de critérios fonológicos, já que se pretendem documentar determinados estruturas segmentais.

Prevê-se, neste exercício, e em caso de necessidade, o recurso a perguntas orientadoras cuja resposta corresponda ao vocábulo a obter.

### 2.2.3. Secção 3

Ao contrário das anteriores, assentes em tarefas de produção textual e lexical, a Secção 3 integra exercícios de leitura (oral), permitindo, portanto, avaliar aspetos criteriosamente selecionados no domínio das competências fonológica e ortoépica, segmental ou prosódica<sup>6</sup>. Por outro lado, a natureza escrita do estímulo permitirá também observar o comportamento dos aprendentes no que diz respeito ao domínio das relações entre unidades gráficas e estruturas fónicas que vigoram na ortografia do português, tanto mais que os dados aqui obtidos podem ser comparados com os elicitados nas tarefas anteriormente descritas.

Assim, a primeira tarefa desta última secção (Tarefa 5) consiste no exercício de leitura oral de um texto, composto por 263 palavras, que o aprendente lê, primeiramente, em silêncio. Pretende-se, com essa estratégia, minorar os efeitos, sobre o ato de leitura, de alguma eventual estranheza no domínio lexical.

---

6 A competência fonológica envolve a capacidade de perceção e produção: (i) das unidades fonológicas e da sua realização em contextos específicos; (ii) dos traços fonéticos que distinguem os fonemas; (iii) da composição fonética das palavras (aqui incluindo a estrutura silábica e o mecanismo prosódico do acento); (iv) da fonética da frase; (v) dos processos fonológicos como a redução vocálica, a assimilação e a elisão (QECRL, 2001: 166). Note-se que a natureza do projeto o limita à componente de produção.



Ao contrário dos exercícios de produção isolada de vocábulos (como o solicitado na Tarefa 4) e de leitura de listas de palavras (correspondendo à tarefa seguinte), a leitura de uma sequência textual com a extensão da utilizada neste Protocolo, torna possível avaliar o modo como o aprendiz domina os processos fonológicos relacionados com a sequência sintagmática e os mecanismos prosódicos relacionados com a diversidade expressiva. Recolhem-se, portanto, dados capazes de apoiar trabalhos no domínio da fonética sintática e da prosódia entoacional, uma vez que, na leitura de texto, se elicit a produção de estruturas não só declarativas, mas também interrogativas e exclamativas.

Com objetivos distintos, a outra tarefa incluída na Secção 3 (Tarefa 6) corresponde a exercícios de *Leitura de listas de palavras*. Num primeiro momento (Tarefa 6.1), o informante é confrontado com um conjunto de 24 vocábulos, onde se incluem 10 pares constituídos por formas simples e formas sufixadas (que integram o radical das primeiras como base) e, ainda, 4 palavras que representam exceções lexicais à regra de elevação vocálica do português europeu contemporâneo (PEC). Avalia-se, portanto, o domínio que os aprendentes revelam, por um lado, das regras de colocação do acento em português e, por outro, do mecanismo da elevação do vocalismo em posição átona, bem como das exceções (regulares ou marcadas) a ele associadas. Imediatamente a seguir (Tarefa 6.2), o informante é confrontado com 49 itens que, de forma não sequencial, permitem constituir diferentes pares mínimos<sup>7</sup>:

Lista de palavras (6.1) - exemplos	Lista de palavras (6.2) - exemplos
cedo	lua
cedinho	rua
fogo	gola
fogão	cola
belo	fado
beleza	fato
fácil	doce
facilmente	dose

<sup>7</sup> Nalguns casos, uma mesma palavra surge associada a duas outras, por via da comutação de segmentos fonológicos diferentes: por exemplo, *vela* opõe-se a *bela*, mas também a *velha*; *vaca* opõe-se a *vaga*, mas também a *faca*. Noutros casos, o informante é confrontado com pares de palavras que, não sendo pares mínimos no sentido fonológico da expressão, permitem igualmente avaliar a interiorização da distintividade de alguns traços em português (cf. *metas* [nome] / *mesas*, par com que se pretende avaliar o conhecimento da oposição entre os timbres vocálicos anteriores médios, ou *calções* / *canções*, com que se pretende avaliar o domínio da oposição entre sons orais e nasais no domínio do vocalismo).



corar	doze (o) peso (o) penso
-------	-------------------------------

Figura 2: Exemplos de léxico motivado na Tarefa 6 (leitura de listas de palavras)

Refira-se, ainda, que, em todas as tarefas da Secção 3, é solicitado ao informante que assinale os itens que não conhece; por via deste procedimento, recolhem-se os dados que permitem avaliar o efeito, na situação de leitura e na qualidade da produção, de eventuais situações de desconhecimento lexical.

### 2.3. Seleção lexical

Uma questão essencial que se coloca no momento da organização do exercício de *Nomeação de figuras* e dos exercícios que constituem a Secção 3 do questionário (todos eles orientados para a recolha de dados relevantes na área da aquisição e desenvolvimento da fonologia de PL2) é a que diz respeito à seleção do léxico aí utilizado. Assim, a escolha do vocabulário a elicitar foi orientada de modo a documentar estruturas, previamente definidas, do sistema segmental da variedade europeia da língua portuguesa:

- i) traços distintivos que estruturam o sistema fonológico, seja no plano do consonantismo, seja no plano do vocalismo: as palavras selecionadas para o exercício de leitura de listas de palavras constituem, em muitos casos, pares mínimos assentes nas distinções articulatórias relacionadas com modo de articulação, ponto de articulação e vozeamento (no caso das consoantes) e com localização, grau de abertura e ressonância nasal, no caso das vogais;
- ii) regras fonológicas contextuais, nomeadamente as relacionadas com o processo de redução do vocalismo em posição átona e com a atualização das unidades sibilante, lateral e vibrante em coda silábica;
- iii) exceções relativas ao mecanismo da redução vocálica, introduzindo-se (nos vários exercícios) vocábulos que constituem casos regulares (como *altura*) ou casos lexicalmente marcados (como *corar*);



iv) formas de realização de estruturas sujeitas a variação alofónica no espaço linguístico português e lusófono, como os ditongos nasais [ẽĩ] e [ẽĩ̃] e os ditongos orais [ow] e [ej].

Definidas as estruturas a testar, e considerando a relação que o grau de familiaridade do vocabulário necessariamente estabelece com a qualidade do desempenho dos informantes, havia que privilegiar léxico conhecido e frequente<sup>8</sup>. Por outro lado, ainda no plano formal, deu-se prioridade às classes nominal e adjetival e a sequências constituídas por duas (ou três) sílabas, atendendo a que a classe das palavras e a sua extensão são, para além da frequência, duas das variáveis com mais relevância para o respetivo processamento psicolinguístico (Festas *et al.*, 2006: 724-726).

Assim, começou por se proceder a uma consulta de manuais escolares concebidos para o ensino de diferentes níveis de PL2<sup>9</sup> e inventariaram-se os adjetivos e nomes que integravam listas de léxico aí compiladas em módulos de sistematização, quer os relativos a áreas léxico-semânticas, quer os relativos a aspetos fonético-fonológicos e ortográficos. Os termos que assim se obtiveram, bem como todos aqueles que foi necessário acrescentar ao inventário inicial que daí resultou, foram, depois, sujeitos a uma análise do grau de frequência, a partir da base fornecida no CORLEX (Nascimento, 2003; [http://www.clul.ul.pt/sectores/linguistica\\_de\\_corpus/projecto\\_lmcp.php](http://www.clul.ul.pt/sectores/linguistica_de_corpus/projecto_lmcp.php)).

Seguindo o procedimento utilizado em Festas *et al.* (2006: 725), foi, então, possível distribuir os vocábulos utilizados por três categorias em função dos respetivos índices de ocorrência. Assim, definiu-se «como “pouco frequente” o limite máximo de 27 ocorrências e como “muito frequente” o limite mínimo de 162 ocorrências».

Em função deste critério, a esmagadora maioria dos vocábulos utilizados nos exercícios em causa pertence à categoria dos lemas mais frequentes<sup>10</sup>, sendo escassos aqueles que constituem

8 «A frequência é uma (...) variável que se sabe ter importância no processamento das palavras. O reconhecimento de palavras é mais fácil e mais rápido quando se trata de palavras muito frequentes, comparativamente a palavras pouco frequentes, tanto no que se refere a palavras ouvidas (Savin, 1963), como no que se refere a palavras lidas (Forster & Chambers, 1973)» (Festas *et al.*, 2006: 725).

9 Concretamente, procedeu-se à análise dos manuais *Aprender Português 1*, *Aprender Português 2* e *Aprender Português 3* e *Português XXI 1*, *Português XXI 2* e *Português XXI 3*.

10 Considerámos, para este efeito, a lematização que encontramos na base do CORLEX (Nascimento, 2003). Aí são apresentadas, sob um mesmo lema, as formas morfológicamente simples e as formas derivadas por sufixação avaliativa. Nesses casos, há uma considerável discrepância entre o número de ocorrências do lema e o número de ocorrências da forma registada, que pode ser substancialmente mais baixo. Considerando, no entanto, que se trata de itens que apresentam sufixos muito produtivos (como -inh(o/a) ou -íssim(o/a)), não nos parece que tal facto ponha em causa a categorização dos dados.



itens pouco frequentes. Estão nesta categoria apenas 3 lemas: *aquecedor* (motivado na Tarefa 4), *cato* (integrado no exercício de leitura de texto) e *estafa* (incluído na lista de leitura apresentada na Tarefa 6.2). Por outro lado, são 20 os lemas (nominais ou adjetivais) cujos índices de ocorrência nos permitem considerá-los frequentes: *cabaz*; *cabeleira*; *cabeleireiro*; *calção*; *chato*; *cola*; *corar*; *desempregado*; *dourado*; *estimação*; *esverdeado*; *fadista*; *feriado*; *Foz*; *gola*; *jato*; *padeiro*; *pastel*; *pomba*; *tarte*.

De modo a poder observar a estabilidade ou a variabilidade do comportamento dos aprendentes, em função do condicionamento contextual, alguns dos vocábulos seleccionados foram usados em mais do que um exercício:

Lema	Tarefa			
	4	5	6.1	6.2
altura		X	X	
belo		X	X	X
cabelo	X	X		
caro				
carro		X		X
casa		X	X	
diretor		X	X	
fado	X			X
hospital		X	X	
mãe	X	X		
padeiro	X		X	
rua		X		X
sozinho	X	X		
vela		X		X

Figura 3: Lemas retomados

### 3. Conclusões

O projeto que aqui se descreveu tem como *output* a constituição de uma base de dados recolhidos junto de aprendentes adultos de PL2 que, em Portugal, aprendem português em contexto instrucional. O *corpus* linguístico que se pretende disponibilizar é oral e, por isso, o acervo em causa destaca-se daqueles que são constituídos por produções escritas ou que resultam



da transcrição de produções orais e permite fundamentar empiricamente estudos em domínios aos quais o investigador não pode aceder por via dos *corpora* já disponíveis. Destaca-se, aí, a investigação nos domínios fónico e pragmático. Por outro lado, e como se foi realçando ao longo deste trabalho, a estrutura do Protocolo e a diversificação dos procedimentos concebidos para a recolha dos dados alarga praticamente a todas as áreas de trabalho o leque de potencialidades dos materiais elicitados.

A recolha de dados sociolinguísticos pormenorizados a que, simultaneamente, se procedeu permite controlar variáveis como a LM e o nível de proficiência linguística de cada sujeito. Estudos sobre fenómenos de transferência e sobre a relação entre níveis formais de aprendizagem e o desenvolvimento das diferentes competências em PL2 podem, então, desenvolver-se sobre uma base robusta de evidências empíricas.

A utilização de uma mesma sequência de estímulos, independentemente do nível frequentado por cada informante e da respetiva LM, assegura a obtenção sistemática das estruturas pré-definidas e permite, portanto, a homogeneidade dos materiais e a comparabilidade dos dados. Tais qualidades compensam a “artificialidade” que se pode criticar a estas formas de obtenção de dados, elicitados e gravados e não espontâneos.

Na verdade, a obtenção de dados com o recurso a inquéritos pré-elaborados constitui uma questão teórica e metodológica há muito equacionada no âmbito dos estudos linguísticos e, particularmente, no âmbito dos estudos de variação. É consensual a ideia de que, por mais estratégias que se desenvolvam para contornar o “paradoxo do observador” (Labov, 1966: 91), nunca se terá a certeza de se obter, por parte do informante, o mesmo enunciado que ele produziria em situação natural de interação.

No domínio concreto da Pragmática, o método aqui utilizado, a elicitación de dados, sobretudo no que toca à produção de atos ilocutórios, pode suscitar algumas reservas, na medida em que fornece dados que podem divergir bastante dos obtidos com a gravação de interações reais (Kasper, 2000: 317-318; Beebe & Cummings, 1996). Todavia, a elicitación de dados tem vantagens evidentes, que já realçámos: permite controlar o contexto e as variáveis que podem influir na produção discursiva e permite obter dados comparáveis, o que nem sempre acontece



com a gravação de interações autênticas, uma vez que, nestas, a variabilidade de atos ilocutórios pode ser diminuta (Billmyer & Varghese, 2000; Bou-Franch & Lorenzo-Dus, 2008).

Assim, globalmente, e apesar das limitações que conscientemente se assumiram, cremos que os materiais obtidos através das diferentes tarefas tornam este *corpus* uma ferramenta inovadora e de amplo espectro no domínio da investigação em PL2.

### Referências:

- Archibald, John (1988) *Second language phonology*. Amsterdam/Philadelphia: John Benjamins.
- Barron, Anne (2003) *Acquisition in interlanguage pragmatics: learning how to do things with words in a study abroad context*. Amsterdam/Philadelphia: John Benjamins.
- Beebe, Leslie & Martha C. Cummings (1996) Natural speech act data versus written questionnaire data: how data collection method affects speech act performance. In Susan Gass & J. Neu (eds.) *Speech Act Across Cultures: Challenges to Communication in a Second Language*. Berlin: Mouton de Gruyter, pp. 65-86.
- Billmyer, Kristine & Manka Varghese (2000) Investigating instrument-based pragmatic variability: Effects of enhancing discourse completion tests". *Applied Linguistics*, 21/4, pp. 517-552.
- Bou-Franch, Patricia & Nuria Lorenzo-Dus (2008) Natural versus elicited data in cross-cultural speech act realization: The case of requests in Peninsular Spanish and British English. *Spanish in Context*, 5(2), pp. 246-277.)
- Corpus de Aquisição de L2 - CAL2. Disponível em: <http://cal2.clunl.edu.pt/>
- Corpus de Produções Escritas de Aprendentes de PL2 - PEAPL2. Disponível em: <http://www.uc.pt/fluc/rcpl2/>.



*Dados de Aprendizagem de Português Língua Estrangeira*. Disponível em:  
<http://www.clul.ul.pt/pt/recursos/314-corpora-of-ple>

Festas, Isabel; J. A. Leitão, M. Dores Formosinho, C. Albuquerque, M. Vilar, C. Martins, A. Branco, L. André, J. Lains, N. Rodrigues, N. Teixeira (2006) PAL-PORT – Uma Bateria de Avaliação Psicolinguística das Afasias e de outras Perturbações da Linguagem para a População Portuguesa. In C. Machado, L. Almeida, A. Guisande, M. Gonçalves & V. Ramalho (eds.) *XI Conferência Internacional Avaliação Psicológica: Formas e Contextos*. Braga: Psiquilibrios, pp. 719-729.

Kasper, Gabriele (2000) Data collection in pragmatics research. In Helen Spencer-Oatey (ed.). *Culturally speaking: Managing rapport through talk across cultures*. London, UK: Continuum, pp. 316-369.

O'Keeffe, Anne & Michael McCarthy (2010) Historical perspective: what are corpora and how have they evolved? In Anne O'Keeffe & Michael McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 3-13.

Labov, William (1966) *The Social Stratification of English in New York City*. Washington: Center for Applied Linguistics.

Martins, Cristina (2013) O corpus de produções escritas de aprendentes de PL2 (PEAPL2/CELGA): caracterização e desenvolvimento de uma infraestrutura de investigação. In R. Bizarro, M. A. Moreira & C. Flores (eds.) *Português Língua Não Materna: Investigação e Ensino*. Lisboa: Lidel, pp. 70-79.

McEnery, Tony & Richard Xiao (2011) What Corpora can offer in Language Teaching and Learning. In: Eli Hinkel (ed.) *Handbook of Research in Second Language Teaching and Learning*, vol. II. New York / London: Routledge, pp. 364-380.

Nascimento, M. Fernanda Bacelar (2002) O lugar do corpus na investigação linguística. In A. Mendes & T. Freitas (orgs.) *Actas do XVIII Encontro da Associação Portuguesa de Linguística*. Porto: APL, pp. 601-605.

Nascimento, M. Fernanda Bacelar (org.), 2003: *CORLEX. Léxico Multifuncional Computorizado do Português Contemporâneo*. Centro de Linguística da Universidade de Lisboa. Disponível



em: <http://www.clul.ul.pt/en/resources/88-project-multifunctional-computational-lexicon-of-contemporary-portuguese-r>

Oliveira, Carla, M. José Ballmann & M. Luísa Coelho (2006) *Aprender Português, 1*. Lisboa: Texto Editores.

Oliveira, Carla & M. Luísa Coelho (2007) *Aprender Português, 2*. Lisboa: Texto Editores.

Oliveira, Carla & M. Luísa Coelho (2007) *Aprender Português, 3*. Lisboa: Texto Editores.

Pereira, Isabel; Cristina Martins; Isabel A. Santos (no prelo) Da tradição à modernidade: ensino, formação e investigação em português L2 na Universidade de Coimbra. In Marta Kfourri (ed.) *Português Língua Estrangeira em Contextos Universitários: experiências de ensino e de formação docente*. Campinas: Editora Mercado de Letras.

*Quadro Europeu Comum de Referência para as Línguas. Aprendizagem, ensino, avaliação* (2001) Conselho da Europa. Lisboa: Asa.

Rio-Torto, Graça (2014) Passado e presente dos *Cursos de Férias*. Da edição de 1924-1925 à de 2014. In Rio-Torto, Graça (coord.) *90 anos de ensino de língua e cultura portuguesas para estrangeiros na Faculdade de Letras da Universidade de Coimbra*. Coimbra: Imprensa da Universidade de Coimbra, pp. 13-38.

Tavares, Ana (2004) *Português XXI, 1*. Lisboa: Lidel.

Tavares, Ana (2005) *Português XXI, 3*. Lisboa: Lidel.

Tavares, Ana (2006) *Português XXI, 3*. Lisboa: Lidel.

Thomas, J. (1983) Cross- cultural pragmatic failure. *Applied Linguistics*, 4 (2), pp. 91-112.

