

**APRESENTAÇÃO PRELIMINAR DO *CPE FACES*:
um 'Corpus de Português Europeu Falado por
Adolescentes em Contexto Escolar'
para o Estudo da Prosódia dos Estilos de Fala***

Ana Isabel Mata
Universidade de Lisboa – F.L.

Introdução

O *corpus* que aqui se apresenta tem apenas alguns meses de existência e dá pelo nome de *CPE FACES*: '*Corpus de Português Europeu Falado por Adolescentes em Contexto Escolar*'. O objectivo principal da sua recolha foi o de reunir um conjunto de textos falados exemplificativos de alguns dos tipos de actividades de produção oral familiares para os alunos no contexto escolar, dentro e fora da sala de aula. Produções orais que fossem entendidas como situações de comunicação autênticas e decorressem com a maior naturalidade possível, cuja gravação permitisse obter a qualidade sonora necessária a uma análise prosódica do Português. Concebido como suporte específico de um estudo de natureza prosódica, tem como principal finalidade ajudar a revelar alguns aspectos da associação e da separação prosódica das sequências de fala no discurso e contribuir para a caracterização de relações de variação prosódica entre estilos de fala.

Este *corpus* reúne um conjunto de textos falados com características situacionais diferenciadas, cuja ocorrência se assume como regular no contexto escolar português, em que os alunos participam no papel de falantes. Textos que podem ilustrar semelhanças e diferenças linguísticas entre estilos de fala comuns na escola, e revelar características da adaptação dos alunos a esses estilos de fala.

Integra gravações áudio e vídeo de 100 textos falados (mais de 10 horas de gravação) por 25 adolescentes dos dois sexos que são naturais de Lisboa e/ou residem na área de Lisboa desde o Ciclo inicial da escolaridade obrigatória. Estes alunos frequentam o último ano da escolaridade obrigatória em duas Escolas Secundárias de Lisboa e fazem parte de duas turmas de Língua Portuguesa leccionadas por professoras que colaboram com a formação de professores da FLUL há já alguns anos.

A primeira fase da constituição do CPE FACES, que implicou a organização de todo o processo de recolha e a gravação das produções orais, a verificação preliminar das gravações e a preparação de cópias para a transcrição, ficou completa em Abril deste ano. A segunda fase de desenvolvimento do *corpus*, envolvendo a preparação do formato e convenções de transcrição, a realização da transcrição ortográfica e da transcrição "miscelânea" – i.e. a anotação de efeitos de fenómenos comuns em actividades de fala ditas "espontâneas" que podem afectar as características do sinal de fala e/ou originar perturbações nas componentes prosódicas (disfluências, riso, tosse, tamborilar com os dedos na mesa, etc.) – dos dados relativos a 60% dos informantes, está praticamente completa¹. A terceira fase, que inclui a análise e transcrição prosódica, e a constituição de uma base de dados com a informação dos múltiplos níveis de transcrição do *corpus*, encontra-se em curso, prevendo-se que fique completa durante o primeiro trimestre de 1996².

Ainda que a sua dimensão ultrapasse as possibilidades da análise e condições inerentes ao projecto específico que agora se desenvolve, o CPE FACES é, indubitavelmente, um *corpus* pequeno com limitações várias (quanto ao número de informantes, aos grupos etários e níveis de ensino, à(s) área(s) geográfica(s) de origem dos falantes que o integram; quanto ao tipo de actividades de produção de fala que o compõe, ...). Creio, no entanto, que os traços que o caracterizam fazem dele um recurso a estimar entre os *corpora* existentes para o Português falado.

Faz-se nesta comunicação uma descrição sumária e preliminar do CPE FACES, incidindo sobre a primeira fase da sua constituição (i.e. a sua recolha), e tecem-se algumas considerações gerais sobre a utilidade futura deste *corpus*³.

Passo então a apresentar alguns traços fundamentais da identidade deste *corpus*.

Descrição do Corpus

Em função da finalidade específica do CPE FACES, que acima introduzi, o processo de recolha dos dados teve em consideração um conjunto de diferentes critérios de base. Fundamentalmente:

- (i) Critério de diversidade das actividades de produção de fala por informante – de modo a captar uma certa gama de variação que permitisse comparar as produções dos alunos em algumas das situações quotidianas habituais dentro e fora da sala de aula;
- (ii) Critério de autenticidade e naturalidade (a maior possível) de ocorrência das actividades de produção de fala gravadas – para captar uma produção prosódica "realista" é fundamental que as actividades correspondam a práticas discursivas correntes e sejam genuínas para os alunos;
- (iii) Critério de qualidade sonora das gravações áudio – quando um trabalho de investigação incide sobre aspectos prosódicos de uma língua a qualidade do registo magnético é condição necessária para uma análise satisfatória (perceptiva e acústica);
- (iv) Critério de restrição para a selecção dos informantes – de modo a controlar o mais possível os factores que possam eventualmente condicionar o desempenho dos alunos e evitar a intromissão de "ruídos" que possam causar dificuldades adicionais à análise dos resultados.

Tipos de Dados

De entre as práticas de produção do oral que têm um lugar explícito nos programas de Língua Portuguesa deste Ciclo⁴ e que haviam sido previstas para os alunos durante os períodos escolares, seleccionaram-se duas actividades de produção diversas, com objectivos de aprendizagem e de ensino específicos: uma de fala interactiva e outra expositiva (não lida). A primeira é não planeada e tem um tipo de ouvinte participante singular: um colega da turma. A segunda é planeada e tem um tipo de ouvinte participante plural (homogéneo e restrito): a turma. Seleccionou-se ainda mais uma actividade de interacção e outra de exposição, ambas não planeadas e sem objectivos de aprendizagem específicos.

Assim, cada informante participa como falante em duas actividades de interacção e duas actividades de exposição diferentes. A saber:

Tipos de actividades de produção de fala dos alunos:

Situações de interacção:

1ª – Diálogos entre pares de alunos sobre episódios da história da sua passagem pela escola ao longo dos últimos anos. É não planeado previamente e corresponderá a uma situação quotidiana habitual sobretudo fora da sala de aula.

2ª – Diálogos entre alunos no decorrer de um trabalho de pares solicitado pela professora para a disciplina de Língua Portuguesa. Tem objectivos de aprendizagem e de ensino específicos. Constitui uma situação de trabalho requerida habitualmente pelas disciplinas escolares.

Situações de exposição:

1ª – Relato de cada aluno aos colegas da turma sobre uma experiência pessoal agradável ou desagradável. É não planeado previamente e corresponderá a uma actividade natural em diversas situações, dentro e fora da sala de aula.

2ª – Exposição oral de um aluno, previamente planeada (mas não lida): constitui a apresentação à turma de um trabalho de pesquisa sobre matéria em estudo na disciplina de Língua Portuguesa, solicitado pela professora e realizado normalmente em grupo. Esta é uma situação de produção com objectivos de aprendizagem e de ensino específicos, habitual dentro da aula.

Optou-se por proceder à gravação, nas quatro situações referidas, das intervenções orais de todos os alunos das duas turmas que não se recusassem a ser gravados, independentemente de poderem ser seleccionados para informantes. Procurou-se com isto prevenir situações que pudessem de algum modo desmotivar a participação voluntária dos alunos e prejudicar a manutenção de um bom ambiente

de trabalho ao longo do processo de recolha dos dados em cada turma. Procurou-se também respeitar o mais possível o ritmo de trabalho da turma, para que as actividades de produção de fala seleccionadas para o estudo fossem entendidas pelos alunos como situações de comunicação autênticas e decorressem com a maior naturalidade possível.

Os informantes deram o seu consentimento explícito para a realização das gravações, apenas sabendo que estas se destinavam a um trabalho académico na área do Português.

Origem dos Dados

A recolha dos dados foi efectuada em duas Escolas Secundárias de Lisboa – escolas do ensino público – durante o último trimestre de 1994 e o primeiro trimestre de 1995.

A fim de obter a qualidade sonora necessária na gravação áudio dos diálogos e das exposições (tendo em conta as condições e o equipamento de gravação disponível) procedeu-se do seguinte modo:

- As situações de fala interactiva foram gravadas por cada par de informantes na Biblioteca da escola, sem a presença da investigadora ou de qualquer outra pessoa. A gravação áudio foi realizada em fita magnética com um gravador UHER 4000 e um microfone duplo (UHER M 640). A gravação vídeo foi realizada em cassette de 8mm com uma câmara SONY CCD-V100E (PAL cor) sobre tripé.
- As situações de fala expositiva foram gravadas por cada informante durante o normal decurso da aula de Língua Portuguesa, i.e. na presença da turma, da professora e da investigadora, com um microfone diferente. Para estas situações cada informante usou um microfone "em colar" (SENNHEISER MD 214 V-3)⁵.

Devo dizer que as gravações vídeo têm como função apenas facilitar a identificação posterior dos informantes e permitir, a outros investigadores, a verificação da autenticidade e naturalidade das diversas situações de intervenção oral dos alunos⁶.

Dimensão do Corpus

A conjugação dos factores 'nível de escolaridade (9º ano)', 'avaliação de nível positivo à disciplina de Língua Portuguesa nos anos anteriores do 3º Ciclo do Ensino Básico', 'não ler o trabalho durante a sua apresentação oral à turma' e 'não apresentar qualquer problema de tipo articulatorio e/ou auditivo' constituiu o critério definitivo para a selecção de informantes. O 9º ano de escolaridade foi seleccionado por ser o último ano de um Ciclo do Ensino Básico correspondente ao término da Escolaridade Obrigatória.

Os informantes são naturais de Lisboa e/ou residem na área de Lisboa desde o 1º Ciclo do Ensino Básico. A maioria encontra-se integrada na mesma turma (constituída por 25 alunos), com a mesma professora de Língua Portuguesa desde o 7º ano (o início do 3º Ciclo do Ensino Básico) e acompanha, desde então, a fase da generalização da Reforma Curricular.

O número total de informantes que, gravados nas quatro situações referidas, reúne estas condições corresponde a 25: (16 numa das escolas e 9 na outra). Destes 14 são do sexo masculino (6 numa escola e 8 na outra) e 11 do feminino (10 numa escola e 1 na outra). As suas idades variam entre 14 e 17 anos (48% dos informantes tem 14 anos).

Procedeu-se ainda, em cada escola, à gravação das produções orais das duas professoras de Língua Portuguesa em quatro situações equivalentes às dos alunos, dialogando elas com a respectiva Directora de Turma nas situações de interacção.

No estado actual do seu desenvolvimento, o *corpus* é, pois, constituído por 100 textos falados por adolescentes (25 de cada situação, mais de 10 horas de gravação) e 8 textos falados pelas respectivas professoras de Português (mais de 4 horas de gravação), o que equivale a um total de aproximadamente 15 horas de gravação.

Para terminar, apenas algumas observações gerais sobre as expectativas futuras para o CPE FACES.

Utilidade do Corpus, Intenções e Sonhos para o Futuro

Não parece difícil pensar em algumas das potencialidades de uso futuro para um *corpus* de discurso contínuo não lido, com diálogos e exposições, recolhido de um modo o mais natural possível⁷.

Ele é um instrumento para a descrição das características prosódicas do Português falado por adolescentes dos dois sexos, do nível final da escolaridade obrigatória, e para o estudo prosódico dos estilos de fala.

Para além do seu contributo para a diferenciação linguística de estilos de fala, a sua finalidade primeira, pode, certamente, ser um suporte adequado para projectos de investigação de natureza diferente, não necessariamente em prosódia – para estudar, por exemplo, aspectos lexicais, sintácticos e discursivos do Português – e um recurso linguístico válido para o desenvolvimento de aplicações tecnológicas várias na nossa língua – por exemplo, para testar e desenvolver sistemas de reconhecimento de fala contínua em Português.

Não havendo actualmente em Português Europeu (tanto quanto sei) outro *corpus* exemplificativo de produções orais de adolescentes do final da escolaridade obrigatória⁸, em situações discursivas correntes na escola, e não esquecendo que a importância do domínio da oralidade na área do ensino e da aprendizagem da língua materna na escola (e, dentro desse domínio, os tipos de discurso) tem sido acentuada nos últimos anos, o CPE FACES pode também ser um recurso linguístico relevante para os professores de Língua Portuguesa. Desde logo como fonte de informação sobre semelhanças e diferenças linguísticas de alguns estilos de fala do quotidiano, ajudando assim o professor nas escolhas necessárias à organização da instrução (relativamente a conteúdos, materiais e actividades, para o treino de utentes competentes da língua falada)⁹.

Devo dizer que a obtenção para os alunos do final do Ensino Secundário e, porque não, também para os alunos do Ensino Superior de material idêntico ao que agora integra o CPE FACES é um projecto que gostaria de concretizar logo que possível. A análise linguística das produções orais (ao longo da escolaridade) de diferentes grupos etários/diferentes níveis de ensino contribuiria para uma caracterização do desenvolvimento e do comando de uma gama de variedades de uso do Português falado, que se assume ser essencial para os alunos no contexto educativo¹⁰.

Continuar a análise e a transcrição prosódica do CPE FACES (de preferência nem sempre a solo!), torná-lo disponível para outros trabalhos de investigação de natureza diferente e mesmo para uma base de dados da Língua Portuguesa são outras das intenções e sonhos para

o futuro¹¹. Lamentaria, é claro, que o tempo e a energia dispendidos (incluindo a resistência à "perfidia" dos equipamentos!) e, sobretudo, a generosidade de todos os que "doaram as suas vozes e faces" para o estudo da língua portuguesa não fossem aproveitados por outras investigações, para as quais este *corpus* possa constituir um recurso específico adequado.

Notas

- * O estudo do CPE FACES integra o projecto de doutoramento apresentado à Faculdade de Letras da Universidade de Lisboa pela autora, projecto que se encontra em fase de desenvolvimento. Dado que os *corpora* linguísticos são um dos tópicos de interesse actual e que ocuparam um lugar de destaque no XI Encontro Nacional da APL, entendeu-se oportuna uma primeira apresentação do CPE FACES, basicamente para divulgar a sua existência, para dizer o que é e qual a sua finalidade. (A apresentação oral foi acompanhada de alguns excertos vídeo do *corpus*.)

Não quero deixar de agradecer às Professoras Doutoradas Maria Helena Mira Mateus e Maria Raquel Delgado Martins, orientadoras do referido projecto de doutoramento, cujo apoio e críticas oportunas têm sido essenciais para o desenvolvimento do CPE FACES.

- 1 Na realização da versão ortográfica manuscrita desta primeira fase de transcrição do CPE FACES (que compreende cerca de 9 horas de gravações) contou-se com a colaboração de vários (ex-)alunos da FLUL.
- 2 O formato de transcrição adaptado para o CPE FACES segue, como referência de base, os níveis de descrição, as convenções e distinções categoriais do sistema de transcrição prosódica ToBI (*Tones and Break Indices*). Sobre este sistema veja-se: Silverman, Kim *et al.* (1992). ToBI: a standard for labelling english prosody. *Proceedings of the 1992 International Conference on Spoken Language Processing 2*: 867-870. Alberta, Canada.
Para a digitalização e análise do sinal de fala utiliza-se o *software* da Sencimetrics para PC: *SpeechStation*. A transcrição do *corpus* será inserida numa base de dados para *Windows*, contendo indicações sobre a localização específica dos enunciados nos ficheiros de fala criados pelo *SpeechStation*.
- 3 Quanto à fase de desenvolvimento do CPE FACES veja-se a breve referência acima incluída. Deixa-se para uma apresentação posterior a descrição e a discussão dos aspectos específicos dos múltiplos níveis de transcrição deste *corpus*.

Definir o formato de transcrição, os traços e tipos de informação a captar, os critérios e as convenções de notação para um *corpus* oral é, como se sabe, uma tarefa complexa e uma (outra) etapa essencial no trabalho de investigação. Sabe-se, também, que isso nunca é (totalmente) neutro em relação aos objectivos da investigação e ao quadro teórico assumido, e pode afectar a percepção das propriedades do discurso oral. Uma discussão de diversas implicações dos sistemas de transcrição do oral é apresentada em: Edwards, Jane A. (1993).

Principles and contrasting systems of discourse transcription. In J. A. Edwards & M. D. Lampert (Eds.), *Talking Data: Transcription and Coding in Discourse Research* (pp. 3-31). Hillsdale, New Jersey: Lawrence Erlbaum Ass.. A discussão das questões relativas à transcrição de um *corpus* oral de Português Europeu é apresentada em: Bacelar do Nascimento, M^a. Fernanda (1987). Um *corpus* de língua falada. In M^a. F. Bacelar do Nascimento, M^a. L. Garcia Marques & M^a. L. Segura da Cruz, *Português Fundamental*, Vol. II, Tomo I (pp. 29-75). Lisboa: INIC/CLUL.

- 4 Ver: DGEBS (1991). *Programa de Língua Portuguesa*. Plano de Organização do Ensino-Aprendizagem, Ensino Básico, 3º Ciclo, Vol. II. Lisboa: Ministério da Educação. Ver ainda: DGEBS (1993). *Materials de Apoio aos Novos Programas*. Língua Portuguesa, 2º e 3º Ciclos, Ficheiros de Jogos, Técnicas e Exercícios, Destinatário: Turma. Lisboa: Ministério da Educação.
 - 5 O equipamento de gravação foi disponibilizado pela FLUL e a sua preparação dependeu dos esforços desenvolvidos pelos funcionários dos Meios Audio-Visuais.
 - 6 Lamentavelmente veio a verificar-se que, devido ao envelhecimento do equipamento, algumas partes das gravações vídeo ficaram danificadas.
 - 7 Que se apresentará transcrito ortográfica e prosodicamente.
 - 8 O *Corpus de Aquisição do Português Europeu* – desenvolvido no Laboratório de Psicolinguística da FLUL sob a orientação de I. Hub Faria – reúne diversos tipos de produções orais de crianças somente até aos doze anos de idade. Ver: Hub Faria, Isabel & Batoréo, Hanna J. (1994). *Corpus de aquisição do português europeu: base de dados CHILDES*. *Revista Internacional de Língua Portuguesa 11*: 137-145. Lisboa: Associação das Universidades de Língua Portuguesa.
 - 9 A indispensabilidade de um conhecimento linguístico fundamentado para o ensino adequado da língua materna é defendida em: Delgado Martins, M^a. Raquel & Duarte, Inês (1993). Brincar com a linguagem, conhecer a língua, fazer gramática. In F. Sequeira (Org.), *Linguagem e Desenvolvimento* (pp. 9-16). Braga: IE - Universidade do Minho. Para o ensino de uma língua estrangeira ver: Delgado Martins, M^a. Raquel (1992). Língua materna/língua estrangeira alguns dados de investigação-acção. *Revue de Phonetique Appliquée: Investigation et Pratique - Vers une Convergence*, 103-104: 119-128. Université de Mons Hainaut.
- Uma argumentação sobre a importância das características prosódicas das variedades de uso da língua falada no ensino e na aprendizagem é apresentada em: Johns-Lewis, Catherine (1986). Conversation as listening material: the prosodic bases of difficulty. *British Studies in Applied Linguistics 1*: 85-94. CILT/British Association for Applied Linguistics. A relevância dos estudos de *corpora* linguísticos para o ensino e a aprendizagem das línguas é discutida em: Kennedy, Graeme (1992). Preferred ways of putting things with implications for language teaching. In J. Svartvik (Ed.), *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82* (pp. 335-373). Berlin: Mouton de Gruyter.
- 10 A importância desta vertente de aprendizagem da língua materna é acentuada em: Crystal, David (1987). *Child Language, Learning and Linguistics: an Overview for the Teaching and Therapeutic Professions*. 2ª Ed.. London: Edward Arnold. Veja-se, ainda, por exemplo: Carter, Ronald (Ed.) (1990). *Knowledge*

about Language and the Curriculum: the Linc Reader. London: Hodder & Stoughton.

A relevância do desenvolvimento das competências do domínio da oralidade no ensino formal do Português (língua materna) é explicitada em: Delgado Martins, M^a. Raquel (1989). Falar, escrever, ler, ouvir. *Revista Internacional de Língua Portuguesa* 2: 59-61. Lisboa: Associação das Universidades de Língua Portuguesa.

- 11 A análise e transcrição dos dados em curso depende dos objectivos e necessidades específicas do presente projecto de investigação. Significa isso que, por enquanto, apenas parte do CPE FACES será transcrita. O *corpus* só estará disponível depois da conclusão do projecto de doutoramento para o qual foi propositadamente recolhido.