

# A classificação das disfluências como mecanismos de (dis)fluência e os seus contextos prosódicos

Helena Moniz<sup>1,2</sup>, Ana Isabel Mata<sup>1</sup> e Isabel Trancoso<sup>2,3</sup>

<sup>1</sup>FLUL/CLUL; <sup>2</sup>L2F/INESC-ID; <sup>3</sup>IST

## Abstract

This work explores prosodic cues of disfluencies using Classification and Regression Tree techniques. We aim at discriminating the most salient linguistic features on a binary classification of disfluencies as fluent vs. disfluent events. Prosodic phrasing and contour shape were essential to perform this task: disfluencies classified as fluent devices were uttered at break indices 3 or 4, and exhibit *plateau* or ascending contours. In a second experiment, prosodic properties were automatically extracted, in order to analyze the global and idiosyncratic prosodic behaviors of the disfluent categories. Filled pauses, complex sequences of disfluencies and prolongations were the most significantly different categories.

**Keywords/Palavras-chave:** disfluencies, fluency classifications, prosody and CART / disfluências, classificação de fluência, prosódia e CART.

## 1. Introdução

As disfluências, isto é, pausas preenchidas, alongamentos, repetições, apagamentos, substituições, inserções, trunções, marcadores de edição, pausas silenciosas atípicas, erros morfossintáticos ou outros eventos concomitantes desempenham um papel estrutural na organização discursiva (Levelt, 1989; Swerts, 1998; Clark & Wasow, 1998; Allwood *et al.*, 1990; Clark & Fox-Tree, 2002). As disfluências têm sido objecto de estudo em diversas áreas, sendo que, paulatinamente, se assiste a uma simbiose de conhecimentos anteriormente cristalizados. Essa simbiose reflecte a consciência de que as disfluências são parte integrante do discurso, visão gradualmente distante da noção de meros erros. Assim sendo, a análise dos padrões regulares dos eventos ditos de disfluência é crucial não só para a linguística geral e a psicolinguística, como para o processamento automático de fala (Nakatani & Hirschberg, 1994; Shriberg, 1994; Liu *et al.* 2006; Tomokiyo *et al.*, 2006).

Sabe-se, com base nos estudos de Fromkin (1971), Heike (1981), Levelt (1989) e Allwood *et al.* (1990), que as ditas “anomalias da fala” não são resíduos erráticos deixados ao acaso quando o falante interrompe a estruturação da sua mensagem, são, acima de tudo, padrões linguísticos com regularidades sistemáticas. Porém, a perspectiva de análise dos referidos eventos é ainda hoje diversificada: no processamento automático de fala (reconhecimento, síntese ou mesmo tradução), alguns estudos defendem a supressão das disfluências para que a mensagem pretendida pelo

falante possa ser o mais limpa e o mais aproximada possível da sua congénere escrita, (Honal & Schultz, 2005; Liu *et al.*, 2006); outros estudos demonstram que as disfluências devem ser transcritas e integradas, para que a mensagem possa ser o retrato fiel do que o falante produziu e possa, assim, ilustrar a especificidade da situação comunicativa, as estratégias dos falantes, na dupla função de emissores e receptores de uma mensagem, ou até mesmo os estados emotivos dos interlocutores e a dinâmica das interacções comunicativas (Benus *et al.* 2006; Adell *et al.*, 2008; Parlikar *et al.*, 2010). Embora as perspectivas sejam distintas, a consciência de que as disfluências têm padrões regulares, obedecem a critérios distribucionais, podem adequar-se ao material linguístico adjacente a diversos níveis de análise linguística e são mecanismos reguladores e estruturantes do planeamento *on-line* tem vindo gradualmente a aumentar e pode hoje ser encarada como transversal às diferentes disciplinas. Mesmo a supressão das disfluências requer toda uma série de processos, como os da sua identificação - isto é, a identificação dos diversos tipos e dos seus padrões regulares -, da sua integração em modelos de língua (contextos e frequências em função de um extenso conjunto de textos de uma língua) e modelos acústicos (correlatos acústicos de qualidade vocálica e consonântica, frequência fundamental -  $f_0$ , energia e duração) para posterior processamento.

Os processos de inclusão/supressão destes eventos na transcrição do contínuo de fala podem, assim, ser associados a conotações distintas: padrões linguísticos com significado *vs.* erros a suprimir, como se de uma interrupção na transmissão contínua da mensagem se tratasse. A posição adoptada no presente trabalho, na linha dos estudos referidos anteriormente, é a de que os eventos ditos de disfluência são reguladores da fala e possuem propriedades linguísticas sistemáticas. Mesmo a literatura crítica (Fromkin 1971; Shattuck-Hufnagel, 1979; Levelt, 1983; 1989; Levelt *et al.*, 1999; *inter alia*) que utiliza “erros” como conceito operativo para compreender o complexo sistema da fala salienta que esses eventos obedecem a restrições linguísticas e, conseqüentemente, exibem padrões regulares:

«It seems quite evident that (...) in the production of speech, it is not true that ‘anything goes’, or that speech performance obeys no rules, or that the errors are totally random and unexplainable. While we may not be able to explain as yet the exact mechanism involved in speech errors, the errors made are not only highly constrained, but provide information about speech performance which non-deviant speech obscures». Fromkin (1971: 48)

No presente trabalho exploram-se pistas prosódicas de fenómenos ditos de disfluência e procura-se validar a assumpção de que o fraseamento prosódico e os contornos entoacionais são cruciais na classificação de juízos de fluência/disfluência. Mais concretamente, pretende-se pesar o contributo das pistas linguísticas determinantes na classificação de diferentes tipos de disfluências como fluentes ou disfluentes, bem como quantificar o contributo das propriedades linguísticas dos referidos eventos, na tentativa de encontrar limiares que diferenciem fluência de disfluência discursivas.

### 1.1. Breve resenha sobre definições de fluência

O termo *fluência* abrange uma série de conceitos associados quer a Língua Materna, quer à aprendizagem de uma Língua Segunda. São exemplificativos disso termos como “proficiência”, “adaptação a diferentes contextos comunicativos”, “mestria”, *inter alia*. O estudo de Fillmore (1979, reed. 2000) define fluência como um conceito satélite com quatro dimensões: i) a temporal, *i.e.*, a manutenção do fluxo discursivo; ii) a sintático-semântica, ou seja, a coerência e coesão lógicas do discurso; iii) a sócio-pragmática, *i.e.*, o uso apropriado da fala em diferentes contextos comunicativos; e a iv) metafórica, relacionada com o uso metalinguístico da língua e com as explorações metafóricas da mesma. A fluência pode, então, ser definida como o uso efectivo da língua, respeitando todos os módulos da gramática.

Para Lennon (2000), a fluência está associada a fluidez discursiva em função de restrições temporais, devidas ao processamento *on-line* da mensagem:

*«A working definition of fluency might be the rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing. This concept of fluency is applicable in principle to both monolinguals and multilinguals, to native speakers and learners»* Lennon (2000: 26)

Para os estudos acima referidos continua, contudo, a não ser evidente se os conceitos de fluência incluem/excluem os mecanismos ditos de disfluência.

O estudo de Wennerstrom (2000) analisa especificamente o contributo das propriedades prosódicas para a caracterização da fluência discursiva. A partir da análise de diálogos informais entre falantes nativos e não-nativos de inglês, verifica que as estratégias de fluência discursiva estão associadas ao fraseamento prosódico e aos tons de constituinte ou de fronteira. Concretizando, os falantes considerados mais fluentes não produzem o seu discurso fragmentando-o palavra prosódica a palavra prosódica, fazem-no antes respeitando a coerência e coesão dos constituintes prosódicos e produzem tons de fronteira que indicam a continuação discursiva.

### 1.2. Estudos realizados para o português europeu

A organização temporal do discurso, especificamente a distribuição sintáctica das pausas silenciosas e das pausas preenchidas, foi primeiramente analisada por Freitas (1990). Com base em dados de leitura e de fala espontânea, a autora salienta que, predizivelmente, as pausas preenchidas são características do discurso espontâneo, o que as torna mecanismos discriminativos do estilo de fala. A distribuição dos dois tipos de pausa obedece a padrões distintivos: i) as pausas preenchidas são produzidas no interior de um sintagma; ii) as pausas silenciosas, nos dados de leitura, são produzidas em localizações hierarquicamente superiores (orações e frases), sendo que nos dados de fala espontânea ocorrem em fronteiras de oração.

O primeiro estudo a apresentar a frequência relativa dos diferentes tipos de disfluências, a sua distribuição, as combinações entre eventos, bem como os padrões

entoacionais e duracionais foi o de Moniz (2006). Estudos posteriores (Moniz *et al.*, 2007; Moniz *et al.*, 2008) detalharam a análise das pausas preenchidas e dos alongamentos. As pausas preenchidas no português europeu, tal como noutras línguas como o inglês ou o francês, correspondem a vogais centrais alongadas [ɛ:], [i:] ou [ə:], uma consoante nasal [m], ou a combinações de vogal com a coda nasal [ɛ:m]. Os alongamentos finais ou pré-finais de material linguístico, tal como em outras línguas, correspondem a pistas para a identificação de fronteiras de constituintes prosódicos (Falé, 1995; Mata, 1999; Frota, 2000). Estes alongamentos não são, todavia, o que está em análise no âmbito do presente trabalho. Entenda-se por alongamentos como mecanismo de disfluência a vocalização prolongada de palavras funcionais maioritariamente monossilábicas, que geralmente sofrem reduções ou mesmo apagamentos (*e.g.*, conjunções e preposições, como [i:], [ki:] ou [di:]). São também consideradas alongamentos as palavras lexicais/funcionais excessivamente alongadas, relativamente à média do falante, produzidas no interior de uma sequência de reformulações ou clarificações.

Diferentes tipos de disfluências ocorrem em diferentes contextos prosódicos. Estudos prosódicos do português europeu (*e.g.*, Viana, 1987; Falé, 1995; Mata, 1999; Frota, 2000; Vigário, 2003) evidenciaram a necessidade de pelo menos dois níveis de fraseamento, constituintes menores e constituintes maiores (IP). Para a distinção entre ambos concorre o facto de o alongamento final pré-fronteira ser substancialmente superior quando se trata de um constituinte maior do que de um menor, indicando assim, tal como apontado por Frota, 2000; Viana *et al.*, 2007, diferentes graus de força para o constituinte IP. Estes dois níveis são anotados com os índices de ruptura 3 para constituintes menores e 4 para constituintes maiores, tal como no sistema ToBI (Silverman, *et al.*, 1992; Viana *et al.*, 2007). O presente trabalho faz uso dos índices referidos, uma vez que, na análise de dados de fala espontânea e especificamente na análise de disfluências, o nível 3 se revelou muito útil para a caracterização do fraseamento das disfluências, em função dos contextos prosódicos adjacentes. Um outro motivo para o uso destes índices é o da manutenção de um sistema de anotação prosódica que permite comparações inter-linguísticas, com o objectivo de avaliar as estratégias utilizadas na produção de mecanismos de disfluência.

Diferentes pausas preenchidas, por exemplo, tendem a ocorrer em diferentes contextos prosódicos (Moniz, 2006; Moniz *et al.*, 2007): i) *aam* é geralmente produzida após um constituinte prosódico maior (por exemplo, para introduzir um novo tópico discursivo); ii) *aa* ocorre com maior frequência após um constituinte prosódico menor (índice 3); iii) *mm* ocorre expressivamente em posição de coda, isto é, no interior de um constituinte (*e.g.*, [kɪm:]). Os alongamentos ocorrem sobretudo entre orações ou entre constituintes menores, tal como acontece com a pausa preenchida *aa*. Os contornos associados às pausas preenchidas são essencialmente contornos *plateau* ou gradualmente descendentes, enquanto que os alongamentos exibem padrões entoacionais mais complexos.

## 2. Corpora

Para a realização do presente trabalho foram utilizados sub-conjuntos do Corpus de Português Europeu Falado por Adolescentes em Contexto EScolar (CPE-FACES), recolhido por Mata (1999), e do *corpus* LECTRA (Trancoso *et al.*, 2008). O primeiro compreende produções de fala preparada e espontânea em contexto de sala de aula (3 professores e 25 alunos), correspondendo a mais de 15h de gravação. A fala preparada (não lida) compreende as exposições orais dos alunos sobre livros propostos nas orientações programáticas da disciplina de Língua Portuguesa, bem como as exposições dos professores sobre conteúdos programáticos da mesma disciplina. Relativamente à fala espontânea, foi solicitado aos falantes que contassem uma experiência agradável ou desagradável na primeira pessoa.

O *corpus* LECTRA foi recolhido no âmbito do projecto homónimo, com o objectivo de transcrever aulas universitárias para aplicações de ensino via *internet*, em especial, para alunos com deficiência auditiva. O *corpus* integra aulas de 7 disciplinas (6 falantes masculinos e um feminino) e tem aproximadamente 75h, sendo que 27h foram já manualmente anotadas.

Observou-se, neste trabalho, uma amostra dos *corpora* (2h de fala do CPE-FACES e 1,5h do LECTRA), para poder comparar estratégias dos falantes relativamente ao uso dos mecanismos de disfluência em diferentes domínios académicos.

### 2.1. Anotação

A anotação multilinear realizada compreende uma fiada de transcrição ortográfica, uma fiada de descrição das disfluências, de acordo com os modelos propostos por Shriberg (1994) e Eklund (2004), e fiadas adicionais para a anotação prosódica das disfluências e dos seus contextos adjacentes (índices de ruptura, contornos entoacionais e reinicialização de  $f_0$ ), bem como para a anotação morfo-sintáctica dessas mesmas unidades de análise (categorias morfológicas, classificação dos tipos de orações e frases)<sup>1</sup>. Foi ainda acrescentada a classificação de juízos de fluência/disfluência. Os juízos foram feitos pela primeira autora e, posteriormente, foram confirmados por duas linguistas, que anotaram um conjunto de exemplos representativos em função da *facilidade de expressão* (Ejzenberg, 2000), a fim de verificar o acerto em relação às classificações inicialmente obtidas. O acerto obtido para esta amostra entre as três linguistas foi de 95%.

A percentagem de disfluências corresponde a 13,2% (1569 disfluências e 11.851 palavras ortográficas) no CPE-FACES, quando contabilizados todos os falantes na sua totalidade. Se atendermos apenas à produção da professora, o total de disfluências corresponde a 7%, sendo que o professor universitário do LECTRA produz 3,2% de disfluências. As percentagens de ambos os falantes estão de acordo com as descritas, *e.g.*, por Fox-Tree (1995), que reporta médias de 6% para diferentes línguas, ou ainda

---

<sup>1</sup> Para uma descrição mais detalhada da anotação multilinear que foi realizada, *vide* Moniz (2006) e Trancoso *et al.* (2008).

por Shriberg (2001), que indica um intervalo de 5% a 10% de disfluências nas conversas entre humanos.

### 3. Resultados da CART

As Árvores de Regressão e Decisão (CARTs do inglês *Classification and Regression Tree Techniques*, Breiman *et al.*, 1984) correspondem a modelos de predição, que são graficamente representados sob a forma de árvore. Cada ramo da árvore compreende uma pergunta e cada folha corresponde a uma partição dos dados e à respectiva resposta ou classificação. Para a realização da CART, utilizou-se o *software* SAS<sup>2</sup>. Os dados utilizados nas experiências com a CART correspondem a subconjuntos de 2h do CPE-FACES e de 1.5h do LECTRA, ambos anotados a vários níveis, como se referiu acima. Os dados foram divididos em treino, validação e teste (divisões de 60%, 20% e 20%, respectivamente).

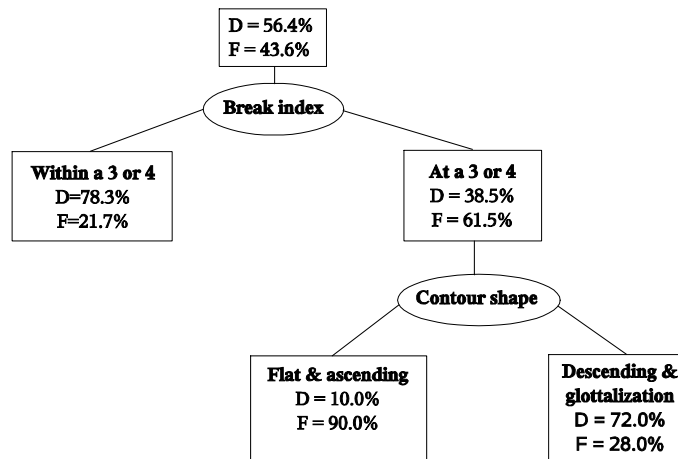


Figura 1: CART com partições acima de 50%

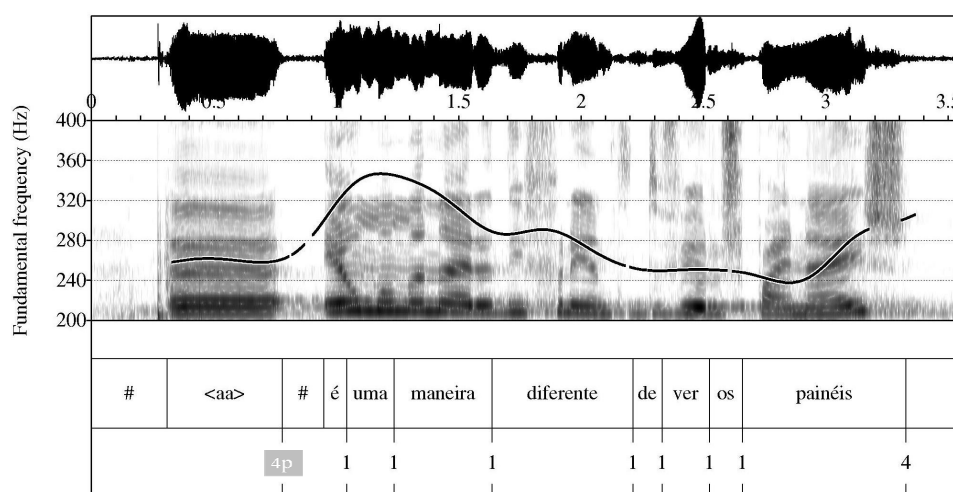
A percentagem de eventos classificados efectivamente como disfluentes é de 56,4%, pelo que 43,6% são classificados como fluentes. Para a discriminação binária fluente/disfluente são tidos em conta todos os níveis da anotação: os juízos de fluência/disfluência (como pista-alvo), a informação prosódica, a informação morfo-sintáctica, os falantes e a situação comunicativa (fala preparada e espontânea).

Os resultados binários de classificação de juízos de fluência vs. disfluência evidenciaram que as pistas prosódicas são a informação mais saliente, possibilitando a diferenciação dos eventos com base em duas partições essenciais na CART: o fraseamento e os contornos prosódicos (ver figura 1).

<sup>2</sup> Ver <http://www.SAS.com>

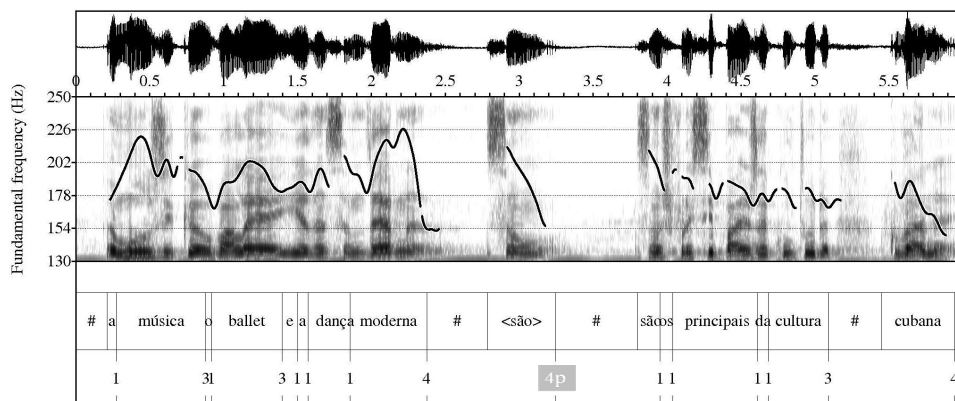
Em relação ao fraseamento, as disfluências produzidas após um constituinte prosódico 3 ou 4<sup>3</sup>, que formam elas próprias um constituinte 3 ou 4, são classificadas maioritariamente como fluentes. Por seu turno, as disfluências produzidas no interior de um constituinte (com índices de ruptura 2p) são predominantemente consideradas disfluentes (78,3%). A percentagem de disfluências produzidas dentro de um constituinte e ainda assim classificadas como fluentes (21,7%) explica-se por dois comportamentos diferentes: i) são eventos produzidos na posição inicial de um constituinte com reinicialização de  $f_0$  (10,4%); ii) são eventos produzidos no final de um constituinte com tons de fronteira que exprimem continuação discursiva, ou ainda no final de um constituinte com contornos de terminalidade, como expectável nas frases declarativas neutras em português europeu.

A segunda partição diz respeito aos contornos entoacionais, evidenciando que os eventos produzidos em fronteiras de constituintes 3 ou 4 com contornos entoacionais *plateau* ou ascendentes são considerados fluentes (90%); os eventos produzidos nas mesmas localizações, mas com contornos descendentes ou com efeitos de glotalização são considerados disfluentes (72%).



**Figura 2:** Exemplo considerado fluente

<sup>3</sup> De acordo com o sistema ToBI, os índices de ruptura correspondem a graus de coesão entre as palavras. São anotados de acordo com uma escala que vai do índice 0 (grau máximo de coesão) ao índice 4. Para o presente trabalho, os índices relevantes são: 2p, 3 e 4. O índice 2p é utilizado para os casos de presença de ruptura temporal (pausas plenas ou virtuais, associadas a alongamentos pronunciados), mas ausência de ruptura tonal. O índice de ruptura 3 corresponde a fronteiras de constituintes prosódicos intermédios (ou menores), não estando associado a casos de terminalidade. Por seu turno, o índice de ruptura 4 corresponde a fronteiras de constituintes entoacionais maiores, caracterizados por fortes rupturas temporais e melódicas.



**Figura 3:** Exemplo considerado disfluyente

As figuras 2 e 3<sup>4</sup> ilustram exemplos classificados como fluentes e disfluentes, respectivamente. A figura 2 mostra a produção de uma pausa preenchida (correspondente a <aa>), que constitui de *per se* um constituinte de índice 4 com contorno *plateau*. Esta pausa preenchida introduz um novo tópico no discurso da professora, funcionando como um suporte para planear e estruturar o tópico que irá ser introduzido. A figura 3 representa uma repetição (o item lexical *são* é repetido). Tal como no exemplo anterior, a repetição forma um constituinte de *per se*, porém com contorno descendente H+L\* L% típico de uma declarativa neutra no português europeu, tal como descrito por Frota (2000). Este padrão entoacional está, neste caso, a ser associado a opções de fraseamento atípicas, promovendo, assim, juízos de disfluência. Contornos *plateau* ou ascendentes (associados a continuação discursiva) são, assim, mais consentâneos com os ajustes necessários à monitorização do discurso, uma vez que indicam coesão entre unidades e, desta forma, se apresentam como estruturalmente mais adequados. No exemplo em análise, a repetição de *são*, ao introduzir uma ruptura no contorno global de  $f_0$  e consequentemente no escalonamento global dos picos de  $f_0$ , pode ser entendida como uma ruptura estruturalmente menos adequada.

Procedeu-se a uma segunda experiência, que consistiu em retirar a pista mais importante, *i.e.*, o fraseamento, e em retrainar a árvore. O objectivo deste treino prende-se com a utilização de pistas que são mais facilmente detectadas de forma automática. Os resultados desta segunda experiência demonstraram que, se os eventos são produzidos com contornos entoacionais *plateau* ou ascendentes e se a informação morfo-sintáctica indica que o constituinte corresponde a um sintagma ou a uma oração, então os eventos são considerados significativamente fluentes (88,7%). Demonstraram também que, se os eventos forem produzidos com reinicialização de  $f_0$ , então são

<sup>4</sup> As figuras foram realizadas com recurso ao *script* de Pauline Welby. Ressalve-se que este *script* pode desenhar contornos de  $f_0$  em partes não vozeadas do sinal.



considerados fluentes em 70,7% dos casos. Caso não se verifique uma reinicialização de  $f_0$  e os contornos sejam descendentes, ou ocorram efeitos de glotalização, os eventos são classificados expressivamente como disfluentes (95,3% e 80%, respectivamente).

Na primeira experiência, o erro na classificação de eventos corresponde a 29,05% e, na segunda, a 32,9%, quando se consideram as 6 folhas<sup>5</sup> mais importantes. Deve salientar-se que em ambas as experiências a classificação das (dis)fluências se encontra acima dos 50%, ou seja, acima do nível impreciso de decisão. Não deixa, contudo, de ser algo surpreendente o facto de a duração dos eventos não ser uma pista expressiva, tal como apontado por Shriberg *et al.* (1997). Somente quando se consideram as 12 primeiras folhas da CART é que a duração parece ter algum peso na classificação, todavia os valores estão abaixo dos 50%.

Os resultados validam a tese de que estes mecanismos estão sob o controlo do falante, demonstrando, também, que existem graus de mestria na produção destes eventos. A granularidade da mestria reflecte-se na maior ou menor integração das disfluências no todo discursivo.

#### **4. Comportamentos globais e idiossincráticos das disfluências**

Na secção anterior verificou-se como o fraseamento e os contornos entoacionais são essenciais para a classificação dos mecanismos de disfluência como fluentes ou disfluentes. Nessa perspectiva, os mecanismos podem mesmo ser classificados como fluentes, quando diversos aspectos suprasegmentais são monitorizados. Pretende-se, na presente secção, caracterizar quantitativamente os parâmetros prosódicos dos mecanismos de (dis)fluência. Mais especificamente, pretende-se verificar se diferentes tipos de disfluências exibem comportamentos distintivos em função dos parâmetros: energia,  $f_0$  e duração. Esta secção tratará, por isso, da caracterização desses parâmetros com vista a uma análise do que podem ser as propriedades globais dos eventos, bem como as suas propriedades idiossincráticas.

Ao subconjunto utilizado anteriormente de 1,5h de fala do corpus LECTRA, acrescentou-se mais 9h extraídas do mesmo *corpus* e anotadas com os mesmos critérios. Estas 10h de fala correspondem a aulas universitárias e foram produzidas por 5 falantes masculinos. Optou-se, nesta fase, por analisar apenas um conjunto de aulas universitárias, para garantir uma maior homogeneidade dos dados. O quadro 1 ilustra a produção de todas as disfluências por falante.

---

<sup>5</sup> Cada folha corresponde a uma partição dos dados. Desta forma, 6 folhas correspondem à sexta partição na árvore.

	Falante1	Falante2	Falante3	Falante4	Falante5	Total
<b>PPs</b>	163	98	589	164	246	<b>1260</b>
<b>COMs</b>	123	70	107	132	214	<b>646</b>
<b>REPs</b>	111	134	68	123	101	<b>502</b>
<b>PRLs</b>	70	32	109	125	101	<b>437</b>
<b>DELs</b>	55	105	18	76	47	<b>301</b>
<b>SUBs</b>	55	53	29	43	36	<b>214</b>
<b>FRAGs</b>	38	34	22	18	36	<b>148</b>
<b>Total</b>	<b>613</b>	<b>526</b>	<b>942</b>	<b>681</b>	<b>746</b>	<b>3508</b>

**Quadro 1:** Produção de disfluências por falante. (“PPs” corresponde a pausas preenchidas, “COMs” a sequências complexas, “REPs” a repetições, “PRLs” a prolongamentos, “DELs” a apagamentos, “SUBs” a substituições e “FRAGs” a fragmentações ou trunicações.)

Como se poderá verificar, a distribuição dos eventos por falante não é idêntica, revelando estratégias diversificadas. De registar que os três tipos de disfluências mais frequentes são as pausas preenchidas, as sequências complexas (com mais do que um tipo de categoria disfluente, *e.g.*, repetição + substituição) e as repetições. Se pausas preenchidas e repetições também noutras línguas são assaz frequentes (Shriberg, 1994; Eklund, 2004; Candeia, 2000), o mesmo não se poderá dizer das sequências complexas. Os resultados obtidos são também diferentes relativamente aos encontrados para um sub-conjunto do *corpus* CPE-FACES (Moniz, 2006), o que poderá indicar estratégias diversificadas não só no que concerne a variável falante, mas também a variável natureza do *corpus* de *per se*.

Os valores mínimos, máximos, médios e medianos de  $f_0$  e de energia, bem como a duração absoluta foram extraídos automaticamente para cada evento disfluente<sup>6</sup>, através da metodologia proposta por Campbell (2007). Os valores de  $f_0$  foram convertidos para a escala de semitons (transformação logarítima da escala em Hertz), utilizando a fórmula proposta por Nolan (2003), sendo que o valor de referência em Hertz corresponde ao mínimo do falante (*e.g.*, mínimo do falante 55Hz, corresponde a 0 na escala de semitons).

Relativamente à questão de se saber se as disfluências no seu conjunto exibem diferenças em relação a cada parâmetro prosódico, o teste Kruskal-Wallis evidencia que existem de facto diferenças significativas ( $p < 0,001$ ) entre os eventos. O resultado obtido suporta a hipótese de que a um nível global as disfluências são prosodicamente distintas, sendo que as pausas preenchidas, as sequências complexas e os alongamentos são estatisticamente significativos em relação a todos os parâmetros prosódicos.

No quadro 2 encontram-se os valores extraídos para os vários tipos de disfluência.

---

<sup>6</sup> A análise dos declives de  $f_0$  está a ser considerada. Num futuro próximo pretende-se disponibilizar os seus resultados.

Cat	$f_0$ (ST)				Energia (dB)				Duração (ms)	
	mediana	méd	máx	Mín	mediana	méd	máx	mín	mediana	méd
<b>FPs</b>	16.7	<b>16.6</b>	<b>19.4</b>	<b>13.4</b>	61.9	<b>61.2</b>	68.5	<b>45.2</b>	360	413
<b>Com</b>	18.7	18.9	<b>22.3</b>	<b>14.3</b>	59.6	<b>59.7</b>	<b>72.5</b>	41.5	600	<b>718</b>
<b>Rep</b>	19.1	18.9	21.3	15.5	58.3	58.2	69.4	42.7	300	383
<b>Prol</b>	17.4	<b>17.8</b>	<b>20.2</b>	<b>14.1</b>	61.3	<b>60.9</b>	70.6	43.7	560	<b>619</b>
<b>Dels</b>	19.6	19.3	<b>22.4</b>	15.4	57.9	57.8	70.4	<b>40.3</b>	360	461
<b>Subs</b>	19.0	18.7	21.0	15.7	57.9	57.9	68.9	43.1	320	343
<b>Frag</b>	19.0	19.1	21.6	16.4	56.9	56.1	68.1	42.6	250	<b>274</b>

**Quadro 2:** A negrito estão representados os valores com níveis de significância de  $p < 0,001$ .

Relativamente aos parâmetros de  $f_0$ , os valores máximos e mínimos de  $f_0$  são mais discriminativos do que os valores médios – estes só permitem a distinção entre pausas preenchidas e alongamentos vs. os restantes eventos. Os apagamentos e as sequências complexas apresentam os valores máximos de  $f_0$ . As pausas preenchidas apresentam valores de  $f_0$  significativamente mais baixos do que todos os outros eventos (tanto no que diz respeito aos valores máximos como aos mínimos de  $f_0$ ). Os alongamentos são significativamente diferentes, uma vez que apresentam valores mais elevados do que os das pausas preenchidas e mais baixos do que os dos restantes eventos.

Quanto ao parâmetro *energia*, as pausas preenchidas são novamente significativamente diferentes, embora não se distingam dos alongamentos ( $z = -.847$ ) no que diz respeito aos valores médios de energia. Acrescente-se que as sequências complexas de disfluências exibem os valores mais elevados de energia.

No que diz respeito ao parâmetro *duração*, as sequências complexas e os alongamentos são os eventos mais longos e significativamente distintos de todos os outros. Os fragmentos, tal como esperado, são os eventos mais breves.

Quando todos os parâmetros são tidos em consideração, dois eventos, porém, não são significativamente diferentes – as repetições e as substituições.

## 5. Conclusões e trabalho futuro

As secções anteriores dão conta das etapas percorridas para uma melhor compreensão do comportamento das disfluências. Os resultados evidenciam que, numa tarefa de classificação de disfluências como mecanismos de fluência/disfluência, as propriedades prosódicas são as pistas linguísticas mais salientes – nomeadamente, o fraseamento prosódico e os contornos da frequência fundamental. Padrões distintos são observáveis e, dado o seu carácter diferenciado, são informativos e discriminativos numa classificação binária desta natureza. Consequentemente, os eventos que são produzidos após uma fronteira de constituinte prosódico 3 ou 4 e que se associam a contornos *plateau* ou ascendentes são expressivamente classificados como fluentes vs.

os que são produzidos no interior de um constituinte, exibindo contornos descendentes ou efeitos de glotalização. Poder-se-á reafirmar, tal como para outras línguas, que os falantes controlam diferentes aspectos segmentais e suprasegmentais e que parecem fazê-lo de uma forma *cirúrgica*, *i.e.*, ajustando-os aos contextos adjacentes. Os resultados suportam a hipótese de que existem diferentes graus de mestria no uso das disfluências e sugerem a possibilidade de estes graus poderem revelar, ou estar associados, a diferentes níveis de conhecimento explícito da língua. Esta é uma hipótese a explorar e a aprofundar em trabalho futuro.

Se na aplicação da CART a importância dos padrões prosódicos foi evidente, procurou-se também verificar se, numa análise mais alargada (cerca de 10h de aulas universitárias), as propriedades prosódicas permitiriam a destriça entre categorias disfluentes. Verificou-se, novamente, a existência de regularidades em função de cada uma das propriedades prosódicas (energia, duração e  $f_0$ ). Quando contrastadas todas as categorias, dois padrões se salientam: i) as pausas preenchidas, as sequências complexas e os alongamentos são as categorias que mais significativamente se distinguem em todos os parâmetros; ii) apenas as repetições e as substituições não se distinguem. Diferentes ilações podem ser retiradas dos resultados obtidos: i) tanto as pausas preenchidas como os alongamentos são vocalizações sustentadas, o que pode explicar o facto de apresentarem características distintivas relativamente a todos os outros eventos; ii) as sequências complexas podem assemelhar-se (pelo menos no seu *onset*) a outros constituintes prosódicos, exibindo os níveis mais elevados de  $f_0$ ; iii) repetições e substituições podem não se distinguir significativamente, por ambas poderem ser associadas a ênfase/reforço informacional.

No trabalho futuro, pretende-se explorar a correlação entre categorias disfluentes e estratégias de reposição de fluência, de modo a destriçar possíveis contrastes ou paralelos com outras estruturas da língua. O objectivo é, em última análise, o de perceber como são monitorizadas as propriedades prosódicas e como podem estas mesmas propriedades contribuir, qualitativa e quantitativamente, para a distinção global entre fluência/disfluência discursiva.

## Referências

- Adell, J., A. Bonafonte & D. Escudero-Mancebo (2008) On the generation of synthetic disfluent speech: local prosodic modifications cased by the insertion of editing terms. In *Proc. Interspeech*, Brisbane, Australia.
- Allwood, J., J. Nivre, & E. Ahls'en, (1990) Speech management - on the non-written life of speech. In *Nordic Journal of Linguistics*, 13.
- Benus, S., F. Enos, J. Hirschberg, & E. Shriberg (2006) Pauses in deceptive speech. In *Speech Prosody Conference*.
- Boersma P. & D. Weenink (2009) *Praat: doing phonetics by computer*. In <http://www.praat.org/>.
- Breiman, L., J. Friedman, R. Olshen & C. Stone (1984) *Classification and Regression Trees*. Pacific Grove, Ca: Wadsworth and Brooks.

- Campbell, N. (2007) Expressive speech and prosody engineering. In F. Chen & C. Jokinen (eds.) *New Trends in Speech Based Interactive Systems*. Springer.
- Candea, M. (2000) *Contribution à l' Etude des Pauses Silencieuses et des Phénomènes dits «d'Hésitation» en Français Oral Spontané – Etude sur un corpus de récit en classe de Français*. Dissertação de Doutoramento, Universidade de Paris III, Sorbonne Nouvelle.
- Clark, H. & T. Wasow (1998) Repeating words in spontaneous speech. In *Cognitive Psychology*, 37, pp. 201-242.
- Clark, H. & J. Fox-Tree (2002) Using uh and um in spontaneous speaking. In *Cognition*, 84, pp. 73-111.
- Eklund, R. (2004) *Disfluency in Swedish Human-Human and Human-Machine Travel Booking Dialogues*. Dissertação de Doutoramento, Universidade de Linköping.
- Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.), *Perspectives on fluency*. Ann Arbor: University of Michigan Press, pp. 287-313.
- Falé, I. (1995) *Fragmento da prosódia do português europeu: as estruturas coordenadas*. Dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Fillmore, C. J. (reed. 2000) On fluency. In H. Riggenbach (ed.) *Perspectives on Fluency*. Universidade de Michigan.
- Fox-Tree, J. (1995) The effects of false starts and repetitions on the subsequent words in spontaneous speech. In *Journal of Memory and Language*, 34, pp. 709-738.
- Freitas, M. J. (1990). *Estratégias de Organização Temporal do Discurso*. Dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Fromkin, V. A. (1971) The non-anomalous nature of anomalous utterances. In *Language*, 47, pp. 27-52.
- Frota, S. (2000) *Prosody and Focus in European Portuguese*. New York & London: Garland Publishing.
- Heike, A. (1981) A content-processing view of hesitation phenomena. In *Language and Speech*, 24, pp.147-160.
- Honal, M. & T. Schultz. (2005). Automatic disfluency removal on recognized spontaneous speech - Rapid adaptation to speaker-dependent disfluencies. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing*. Philadelphia.
- Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (ed.) *Perspectives on Fluency*. Universidade de Michigan, pp. 25-42.
- Levelt, W., A. Roelofs & A. Meyer (1999) A theory of lexical access in speech production. In *Brain and Behavioral Sciences*, 22, pp. 1-38.
- Levelt, W. (1989) *Speaking*. Cambridge, Massachusetts: MIT Press.
- Levelt, W. (1983) Monitoring and self-repair in speech. In *Cognition*, 14, pp. 41-104.
- Liu, Y., E. Shriberg, A., Stolcke, D., Hillard, M., Ostendorf, M., Harper (2006) Enriching speech recognition with automatic detection of sentence boundaries and disfluencies. *IEEE Transaction on Audio, Speech, and Language Processing*.

- Mata, A. I. (1999) *Para o Estudo da Entoação em Fala Espontânea e Preparada no Português Europeu: Metodologias, Resultados e Implicações Didáticas*. Dissertação de Doutoramento, FLUL.
- Moniz, H. (2006) *Contributo para a Caracterização dos Mecanismos de (Dis)Fluência em Português Europeu*. Dissertação de Mestrado, Universidade de Lisboa.
- Moniz, H., A. I. Mata & M. C. Viana (2007) On filled pauses and prolongations in European Portuguese. In *Interspeech 2007*, ISCA, Antuérpia, Bélgica.
- Moniz, H., A. I. Mata, I. Trancoso & M. C. Viana (2008) How can you use disfluencies and still sound as a good speaker?. In *Interspeech 2008*, ISCA, Brisbane, Austrália.
- Nakatani, C. & J. Hirschberg (1994) A corpus-based study of repair cues in spontaneous speech. *Journal Acoustic Society of America*, 95, pp. 1603-1616.
- Nolan, F. (2003) Intonational equivalence: an experimental evaluation of pitch scales. In *15<sup>th</sup> ICPHS*, Barcelona, pp. 771-774.
- Parlikar, A., A. Black & S. Vogel (2010) Improving Speech Synthesis of Machine Translation Output. In *Interspeech 2010*, ISCA Makuhari, Japan.
- Shattuck-Hufnagel, S. (1979) Speech errors as evidence for a serial-ordering mechanism in sentence production. Cooper & Walker (editors), *Sentence processing: Psycholinguistics studies presented to Merrill Garrett*. Laurence Erlbaum.
- Shriberg, E. (1994) *Preliminaries to a Theory of Speech Disfluencies*. Dissertação de Doutoramento, Universidade da Califórnia.
- Shriberg, E., R. A. Bates & A. Stolcke (1997) A prosody-only decision-tree model for disfluency detection. In *Proceedings of Eurospeech*, Rodes, Grécia, pp. 2383-2386.
- Shriberg, E. (2001) To “errrr” is human: ecology and acoustics of speech disfluencies. In *Journal of the International Phonetic Association*, 31, pp. 153-169.
- Silverman, K., M. Beckam, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert & J. Hirschberg (1992), “ToBI: a standard for labeling English prosody”. In *Proceedings ICSLP*, Banff, vol. 2, pp. 867-870.
- Swerts, M. (1998) Filled pauses as markers of discourse structure. In *Journal of Pragmatics*, 30, pp. 485-496.
- Tomokiyo, L., K. Peterson, A., Black, K., Lenzo (2006) Intelligibility of machine translation output in speech synthesis. In *Proceedings of Interspeech*, Pittsburgh.
- Trancoso, I., R. Martins, H. Moniz, A. I. Mata & M. C. Viana (2008). The LECTRA corpus – classroom lecture transcriptions in European Portuguese. In *LREC 2008*, Marrocos.
- Viana, M. C. (1987) *Para a Síntese da Entoação do Português*. Dissertação de Doutoramento, FLUL.
- Viana, M. C., S. Frota, I. Falé, F. Fernandes, I. Mascarenhas, A. I. Mata, H. Moniz & M. Vigário (2007). Towards a P\_ToBI. *PAPI2007. Workshop on the Transcription of Intonation in Ibero-Romance*. Universidade do Minho. (ver <http://www.ling.ohio-state.edu/~tobi/>)
- Vigário, M. (2003) *The prosodic word in European Portuguese*. Mouton de Gruyter, Berlim/ Nova Iorque.
- Wennerstrom, A. (2000) The role of intonation in second language fluency. In H. Riggenbach (ed.) *Perspectives on Fluency*. Universidade de Michigan, pp. 102-127.