

Prosódia, variação e processamento automático

Neste capítulo apresentamos um olhar panorâmico sobre a variação prosódica e sobre a sua interface com a área do processamento automático de fala. Tendo por base essencialmente a investigação que tem sido desenvolvida no português europeu sobre corpora de fala espontânea e preparada, em contextos de exposição e de interação, nomeadamente na variedade padrão falada em Lisboa, analisamos a variação da entoação em contextos declarativos e interrogativos, e abordamos as funções pragmático-discursivas que podem associar-se também a outros parâmetros prosódicos. Partindo de estudos comparativos inter-estilos (com maior/menor grau de espontaneidade e de planeamento, e natureza mais interativa/expositiva) e inter-falantes (espaço geográfico, género, grupo etário/estatuto), destacamos o papel da variação estilística e sociolinguística da prosódia no português europeu. Mostramos também o papel da variação no processamento automático de proeminência prosódica, pontuação, disfluências e emoções.

Palavras-chave: prosódia, entoação, variação estilística, variação social, processamento automático de fala

1. Introdução

São múltiplas as funções desempenhadas pela prosódia. A variação da entoação é um dos mais importantes mecanismos para o agrupamento de constituintes e a demarcação de fronteiras no contínuo de discurso, para a marcação de ênfase e contraste, para a distinção entre intenções comunicativas, e para a expressão de diferenças de atitude, emoção e expressividade. Para além da marcação de natureza melódica, baseada principalmente na variação dos contornos de frequência fundamental (f_0), fazem ainda parte dos mecanismos prosódicos de que os falantes dispõem os parâmetros de marcação temporal (incluindo as pausas) e a variação da energia e da qualidade da voz. Por exemplo, os enunciados declarativos são marcados por uma descida final de f_0 sobre a última sílaba acentuada, por um alongamento do núcleo dessa sílaba, muitas vezes por efeitos de laringalização pré-fronteira, podendo ainda ser seguidos por pausa. Para distinguir terminalidade de continuação, i.e. para sugerir a continuação desses enunciados, pode ser usado um padrão de f_0 diferente, nomeadamente uma subida final de f_0 . Em caso de hesitação, por exemplo antes da introdução de um tópico novo no discurso, a reposição da fluência é marcada nesses enunciados por um contraste prosódico, indicado por diferenças de f_0 (tipo de padrão entoacional, nível e gama de variação mais elevados), mas também por uma maior energia, por uma pausa de maior duração antes do item a destacar e ainda pela menor duração desse item, por comparação com a sua produção disfluente. O controlo da gama de f_0 e de energia e o da velocidade de elocução podem ainda traduzir-se na nítida impressão de uma maior expressividade dos falantes.

Para além de correlações entre variação prosódica e aspetos de outros níveis de informação linguística, nomeadamente sintáticos, semânticos e pragmático-discursivos, a investigação sobre corpora de fala espontânea tem mostrado que, a par dos índices de variação entoacional, também outros parâmetros prosódicos podem ser usados de forma regular

pelos falantes na organização do discurso, contribuindo para a sua fluência comunicativa. A comparação inter-corpora (relatos e diálogos espontâneos, apresentações orais escolares, aulas universitárias e noticiários televisivos), intra-corpora (relatos espontâneos vs. apresentações preparadas não-lidas) e inter-falantes (género, grupo etário, estatuto) tem também mostrado que a variação prosódica configura diferentes estratégias de adaptação dos falantes a estilos de fala distintos e desempenha ainda uma função sociolinguística. Por outro lado, os resultados destes estudos comparativos têm vindo a contribuir para avanços no processamento automático de fala, em especial para a predição de proeminência prosódica, sinais de pontuação, disfluências e marcadores discursivos, ou mesmo para a deteção de eventos paralinguísticos (e.g. emoções).

Neste capítulo incidimos sobre a variação da entoação no português europeu, referindo sempre que oportuno a variação de outros parâmetros, uma vez que não cabe nos limites deste capítulo um tratamento exaustivo das diferentes dimensões prosódicas. Apresentamos também um olhar panorâmico sobre os mais recentes desenvolvimentos na área do processamento automático da prosódia do português europeu. Temos como referência essencialmente os dados acumulados nas últimas décadas a partir da análise de corpora.

Na próxima secção (secção 2), descrevemos o sistema de representação e o modelo teórico adotado para a análise da entoação do português europeu. Ilustramos ainda nesta secção os principais padrões e funções da entoação em contextos declarativos e interrogativos, que têm sido os mais estudados na variedade padrão falada em Lisboa e os mais explorados na área de processamento automático de fala. Na secção 3 introduzimos a variação prosódica numa perspetiva geolinguística, social e estilística. Na secção 4 descrevemos algumas das suas aplicações atuais ao processamento automático da prosódia. Na secção 5 destacamos as principais conclusões e apontamos linhas de trabalho futuro.

2. Sistema de representação e modelo adotado

Para representar a variação dos contornos de f_0 relevante para a distinção quer de padrões quer de constituintes entoacionais no contínuo de fala, há que usar um sistema de transcrição. Preferencialmente, um sistema que: (i) permita especificar diferentes tipos de informação prosódica, sincronizados com o sinal acústico, em linhas de notação distintas, refletindo uma perspetiva modular acerca das componentes da estrutura prosódica e da sua relação com outros níveis de análise linguística; (ii) tenha sido testado com sucesso no estudo da prosódia a partir de *corpora*; e (iii) se fundamente num modelo entoacional considerado consistente e de larga aplicação, uma vez que o ato de transcrever a entoação não é completamente neutro em relação a um modelo teórico, e implica uma interpretação linguística da variação fonética.

Neste capítulo seguimos a adaptação feita para o português (Viana et al. 2007) do sistema ToBI – *Tones and Break Indices*, inicialmente proposto para o inglês americano (Silverman et al. 1992; Pitrelli/Beckman/Hirschberg 1994). O ToBI baseia-se num modelo de interpretação das curvas de f_0 como sequências de tons no quadro da fonologia métrica autosegmental. Desenvolvido essencialmente a partir dos trabalhos de Bruce (1977) para o

sueco, de Pierrehumbert (1980) para o inglês americano, de Beckman e Pierrehumbert (1986) para o inglês americano e o japonês, este modelo é, nas últimas décadas, um dos mais influentes no estudo da entoação das línguas (para uma discussão do modelo, cf. Ladd 1996). É neste quadro geral que têm sido realizados estudos comparativos dos padrões entoacionais de línguas diferentes (e.g., Avesani/Hirschberg/Prieto 1995 para o inglês, o espanhol e o italiano; Grice et al. 1995 para o alemão, o Italiano e o búlgaro; Grabe 1998 para o alemão e o inglês; Frota et al. 2007 para o português europeu, o italiano, o espanhol e o catalão; Jun 2005 para uma comparação inter-línguas mais alargada) e de diferentes variedades da mesma língua (e.g., Grabe et al. 2000, Clopper/Smiljanic 2011 para variedades do inglês britânico e americano, respetivamente; Frota et al. 2015, para variedades do português europeu e do português brasileiro), e ainda muitos dos trabalhos sobre o português europeu nas últimas décadas (e.g. Viana 1987, Falé 1995; Grønnum/Viana 1999; Frota 2000; Vigário 2003; Cabarrão 2013; Moniz 2013; Cruz 2013; Moniz et al. 2014a; Mata et al. 2014a; Mata/Moniz/Batista no prelo; Moniz et al. no prelo).

A simplicidade do esquema de transcrição e o grau de consistência alcançado entre transcritores em diferentes tipos de fala, segundo algumas avaliações já publicadas (cf. Silverman et al. 1992; Pitrelli/Beckman/Hirschberg 1994, para o inglês; Escudero et al. 2012, para o catalão, e referências aí citadas para outras línguas; Mata et al. 2014b, para o português europeu), são ainda razões para, tal como tem acontecido em outras línguas (como por exemplo o G-ToBI para o alemão, ou o Cat-ToBI para o Catalão), se optar, desde a década de noventa, pela adaptação do sistema ToBI na descrição da variação prosódica a partir de corpora de fala do português europeu (cf. projeto CORAL)¹.

O ToBI é constituído por 4 fiadas de transcrição paralelas, alinhadas entre si e com o sinal acústico: (i) fiada ortográfica, (ii) tonal, (iii) de índices de rutura e (iv) miscelânea (fiada opcional, usada frequentemente para registar eventos paralinguísticos, como risos, tosse, disfluências, *inter alia*). Outras fiadas podem ser acrescentadas, por exemplo para a notação de categorias sintático-discursivas cuja análise possa ser considerada relevante para uma interpretação da variação entoacional (cf. Mata et al. 2014a). Uma transcrição multilinear ToBI é ilustrada para o português na figura 1, feita com o programa Praat (Boersma/Weenink 2013). Identifica-se na legenda da figura (entre parêntesis) o corpus de onde foi extraído o exemplo. (O mesmo será feito para os restantes exemplos apresentados neste capítulo.)

¹ https://www.l2f.inesc-id.pt/wiki/index.php/CORAL_-_Labelled_Spoken_Dialogue_Corpus#Publications

Figura 1. Exemplo de uma transcrição multilinear ToBI para o enunciado *tenho, barracas sujas, tenho*. (CORAL)²

As fiadas tonal e de índices de rutura representam o núcleo de uma análise prosódica. Destinam-se à codificação de dois tipos de informação: os tons que compõem o padrão entoacional e o agrupamento de palavras em constituintes prosódicos, respetivamente. A tipologia de tons do ToBI assume dois níveis básicos, tom alto (H) e tom baixo (L), e duas grandes categorias de descritores: (i) os acentos tonais, que atribuem proeminência aos itens lexicais e estão associados a sílabas metricamente acentuadas no enunciado (marcados com ‘*’); (ii) os tons associados a fronteiras de constituintes, que demarcam fronteiras de dois níveis de fraseamento entoacional (marcados com ‘-’ e ‘%’).

Tons

No português estão identificados sete acentos tonais (H+L*, H*+L, L*+H, L+H*, H*, L*, H+!H*) e cinco tons fronteira (L%, H%, !H%, LH%, HL%) – veja-se a representação esquemática de contornos de f_0 na figura 2. Os acentos tonais podem corresponder a um tom simples (L*, H*) ou a eventos bitonais (e.g., L+H*, L*+H, H+!H*), em que o diacrítico “*” distingue o tom que se encontra alinhado com a sílaba acentuada e o diacrítico “!” é utilizado quando há compressão da gama de variação dos níveis altos (!H). O acento tonal que ocorre em posição final de constituinte é, em regra, o mais importante. Por isso, é identificado como nuclear e em conjunto com o tom fronteira forma o designado contorno nuclear.

Figura 2. Esquematização de contornos de f_0 para acentos tonais e tons fronteira, seguindo *Towards a P_ToBI* (Viana et al. 2007).

As linhas — indicam a sílaba acentuada.

Para o português têm sido considerados dois níveis de fraseamento entoacional, com configurações de fronteira de força diferenciada, tanto em termos de duração e extensão do movimento final de f_0 , como em termos da distribuição de pausas e da ocorrência de vários fenómenos de sândi. Frota (2000) e Viana et al. (2007) associaram os dois níveis a uma unidade do tipo constituinte entoacional (*intonational phrase*), distinguindo um constituinte maior (IP) e um constituinte menor (ip), na sequência de Ladd (1996). Para dar conta das diferenças entre estas fronteiras de fraseamento entoacional, usam-se os diacríticos tonais de fronteira ‘%’ e ‘-’, bem como os correspondentes índices de rutura 4 e 3, como veremos mais adiante. (Veja-se Frota 2014 para uma proposta de uso do diacrítico ‘%’ a representar fonologicamente tanto ‘-’ como ‘%’.) O constituinte entoacional menor (ip), marcado com o diacrítico ‘-’, compreende pelo menos um acento tonal e um tom fronteira (L-, H-, e !H-, menos frequentemente LH- e HL-). O constituinte entoacional maior (IP), por seu turno, é constituído por um ou

² O corpus CORAL é constituído por diálogos espontâneos de tipo *map-task* (i.e., diálogos para a indicação de percursos em mapas) entre jovens adultos (Trancoso et al. 1998; ISLRN 499-311-025-331-2).

mais constituintes entoacionais menores e o seu tom fronteira marcado com o diacrítico ‘%’, usado para descrever o movimento final de f_0 (frequentemente L% e H%, que em geral se associam à expressão de terminalidade/não-terminalidade, respetivamente).

Ilustramos em seguida, em contextos declarativos e interrogativos, os principais contornos nucleares e respetivas funções na variedade padrão falada em Lisboa.

H+L* (modelizado como uma descida pronunciada de f_0 que atinge um alvo tonal baixo dentro da sílaba acentuada) é um acento comum em posição nuclear em português europeu. Seguido de um tom fronteira L%, define o contorno característico das declarativas, bem como o das interrogativas Qu-, como os exemplos das figuras 3 (*à tua esquerda.*) e 4 (*onde é que está?*) ilustram. Seguido de um tom fronteira LH%, como no exemplo da figura 5 (*está à tua esquerda?*), define o contorno característico das interrogativas de *sim-não* que funcionam como pedidos de informação (Frota 2002, 2014; Viana et al. 2007; Mata/Santos 2010). Sabe-se que a perceção destas interrogativas depende sobretudo da extensão da subida final de f_0 : as fronteiras ascendentes LH% devem ser superiores a dois semitons (Falé 2005; Falé/Faria 2006) – no que se distinguem das fronteiras LH% que podem surgir em contornos intermédios, em declarativas longas com mais de uma fronteira entoacional. Quanto às interrogativas Qu-, tem sido apontado que estas tendem a distinguir-se do contorno das declarativas por apresentarem, comparativamente, um nível de f_0 inicial mais elevado e uma maior gama de variação (Viana 1987; Falé 2005).

Figura 3. H+L* L% numa declarativa (CORAL)

Figura 4. H+L* L% numa interrogativa Qu- (CORAL)

Figura 5. H+L* LH% numa interrogativa de *sim-não* que solicita informação (CORAL)

L* pode ocorrer como acento nuclear em constituintes não finais no interior de frases complexas, em materiais de leitura (Falé 1995). Em diálogos espontâneos, L* L% é, a par de H+L* L%, um padrão de resposta afirmativa comum para a expressão de concordância com o que foi dito, ou de confirmação (Cabarrão 2013), e.g. *sim*, nas figuras 6 e 7. L* H% é comum para a sugestão de continuação (Viana et al. 2007; Cabarrão 2013), e.g. *sim* na figura 8, uma resposta afirmativa com função reguladora: o

falante indica ao interlocutor que compreendeu a mensagem e quer que este continue a dar-lhe informações/instruções.

| | | |
|-------------------------|--|--|
| A: Que é logo a seguir? | A: O pasto bravio fica-te à tua esquerda e os poços secos à tua direita. | A: Em cima, mais ou menos ali relativamente perto, ficam-te as barracas sujas. |
| B: Sim. | B: Sim. | B: Sim. |

Figura 6. H+L* L% numa resposta afirmativa de confirmação (CORAL) Figura 7. L* L% numa resposta afirmativa de concordância (CORAL) Figura 8. L* H% numa resposta afirmativa com função reguladora (CORAL)

H* e L+H* (modelizado como uma subida pronunciada de f_0 que atinge um alvo tonal alto dentro da sílaba acentuada) são comuns em posição nuclear (Viana et al. 2007), sobretudo seguidos de fronteiras H-/H% (Mata/Moniz/Batista no prelo). Em contextos declarativos, H* surge associado à expressão de valores pragmáticos de continuação em constituintes entoacionais menores, como no exemplo da figura 9 (*e então depois, quando fui para casa.*). Como a figura também ilustra, o mesmo pode acontecer em constituintes entoacionais maiores.

Figura 9. H* H-/H% em contexto declarativo (CPE-FACES)³

Viana et al. (2007) sugeriram que L+H* pode ser usado pelos falantes para assinalar informação nova ou para destacar informação dada que precisa de ser (re)ativada, como no exemplo da figura 10, onde o falante introduz uma lista das qualidades mais apreciadas num animal (*dócil, elegante, amigo*), de acordo com as respostas a um inquérito que aplicou.

Figura 10. L+H* !H-/!H% em contexto declarativo (CPE-FACES)

De facto, em contextos interrogativos (Mata/Santos 2010), L+H* e H* seguidos de fronteira tonal não descendente permitem distinguir interrogativas que solicitam confirmação de perceção (e.g. *por baixo do pátio de juizes?*, na figura 11) vs. interrogativas que funcionam como pedidos de informação (cf. H+L* LH%, em *está à tua esquerda?*, na figura 5).

³ O corpus CPE-FACES é constituído por produções de fala espontânea e preparada (não-lida), de adultos e adolescentes de ambos os sexos, em contexto de sala de aula (cf. Mata 1999; Mata et al. 2014b).

Concorre ainda para a desambiguação entre atos de fala o acento L^*+H (modelizado como um alvo tonal L dentro da sílaba acentuada seguido de uma subida pronunciada de f_0 que normalmente termina na sílaba seguinte). Em contextos interrogativos, associa-se frequentemente a afirmativas de compreensão não neutras, que sugerem incorreção, incredulidade ou surpresa (Mata/Santos 2010). Veja-se o exemplo da figura 12 (*vale irrigado?!).* Além disso, comparativamente às afirmativas de percepção, as afirmativas de compreensão são produzidas num nível de f_0 mais baixo (acento nuclear e tom fronteira).

A: Portanto passas por baixo do pátio dos juizes,
B: Por baixo do pátio de juizes?

Figura 11. $L+H^*$ H% em interrogativa afirmativa de percepção (CORAL)

A: Passas ao vale irrigado.
B: Vale irrigado?! (Eu não tenho vale irrigado!)

Figura 12. L^*+H H% em interrogativa afirmativa de compreensão não neutra (CORAL)

L^*+H é ainda comum em interrogativas *tag* (e.g., *não é?*, na figura 13). Este acento tem também sido associado à marcação de foco prosódico em enunciados interrogativos (Frota 2002).

Figura 13. L^*+H H% em interrogativa *tag* (CORAL)

H^*+L (modelizado como um alvo tonal H dentro da sílaba acentuada seguido de uma descida pronunciada de f_0 que normalmente termina na sílaba seguinte) tem sido associado a estratégias de demarcação prosódica de foco em enunciados declarativos (Frota 2000, 2014), como no exemplo da figura 14, onde *em pedra* é acentuado com H^*+L para assinalar que se trata de um paredão em pedra e não de um paredão de outro tipo.

Figura 14. H*+L L% na marcação de foco prosódico (CORAL)

Quanto a H+!H* (modelizado como um alvo tonal !H dentro da sílaba acentuada que é precedido por um pico de f_0 na sílaba átona anterior), por ser pouco frequente em posição nuclear, está ainda em larga medida por estudar a sua distribuição e o tipo de valores pragmáticos a que pode associar-se. Ilustra-se a sua ocorrência em contexto de reformulação (e.g., *sabia vs. sei*, na figura 15), um dos contextos em que aparece em fala espontânea.

Figura 15. H+!H* H% em contexto de reformulação (CPE-FACES)

Índices de rutura

«Para complementar a informação tonal com uma representação da estrutura rítmica da fala, marca-se na fiada de índices de rutura a força de coerência ou de disjunção entre todas as palavras adjacentes» (Silverman et al. 1992, 868 – tradução nossa). Utiliza-se, para tal, uma escala de índices de rutura, inspirada no trabalho de Price et al. (1991). Esta escala varia entre os níveis 0 e 4. Quanto menor for o valor do índice de rutura, maior será o grau de ligação percebido entre uma palavra e o que lhe está adjacente; quanto maior o valor do índice de rutura, maior o grau de separação percebido nessa fronteira. O índice de rutura 0 indica uma estratégia de conexão máxima entre duas palavras consecutivas, i. e., em português europeu pode corresponder a uma sequência fónica como [’lOg s ’ve]⁴, com a elisão das vogais átonas [u] e [@], em vez de [’lOgu s@ ’ve], correspondente à pronúncia das palavras em forma de citação; ou a [b6’Rak6 ’SuZ6S], com a realização de apenas uma ([S]) das duas fricativas na concatenação das palavras [b6’Rak6S ’suZ6S], como se observa na figura 1. O índice de rutura 1 é o mais comum entre duas palavras consecutivas no interior de um constituinte entoacional, não implicando omissão de segmentos fonéticos (e.g., [kj ’E], forma de sândi externo predizível, em vez de [k@ ’E]). O índice de rutura 2 é utilizado quando o anotador tem dúvidas em relação ao nível a atribuir, motivadas por: i) perçetivamente considerar que se pode tratar de um nível 1, mas a palavra contém pistas tonais ou alongamentos de segmentos, como os que são comuns em pré-fronteira de constituinte entoacional; ou ii) perçetivamente considerar que se pode tratar de um nível 3 ou 4, com pausa real ou virtual, mas ausência de pistas tonais de fronteira de constituinte. Os índices de rutura 3 e 4 correspondem a fronteiras de constituintes entoacionais menores e maiores, respetivamente. De acordo com estudos anteriores sobre corpora de fala espontânea em português europeu (Mata/Moniz/Batista 2014; no prelo; Moniz/Trancoso/Mata 2010), em fronteiras de nível 3 (ip) verifica-se uma frequência mais elevada de tons altos/ascendentes e pausas silenciosas

⁴ Transcrição fonética seguindo o SAMPA (Wells 1997).

subsequentes mais curtas; em fronteiras de nível 4 (IP), uma frequência mais elevada de tons baixos/descendentes e de efeitos fonéticos pré-fronteira (e.g., efeitos de voz laringalizada), e pausas silenciosas subsequentes mais longas. São também pistas prosódicas para a distinção entre estes dois níveis os declives e desvios padrão de f_0 na última palavra do constituinte: os valores são sempre mais elevados para os constituintes maiores do que para os constituintes menores, tanto no caso de tons altos/ascendentes como no caso de tons baixos/descendentes. Para além disso, em transcrições pontuadas, a quantidade de sinais de pontuação é significativamente mais elevada em fronteiras de nível 4 do que em fronteiras de nível 3. Comparando, por exemplo, a distribuição de pontos e de vírgulas, verifica-se que os pontos são associados sobretudo a fronteiras 4, as vírgulas, embora possam aparecer em ambos os níveis, associam-se maioritariamente a fronteiras de nível 3.

Outros tipos de informação prosódica relevantes para a deteção na fala de estruturas de nível hierárquico superior (como constituintes similares a frase, enunciados entoacionais e parágrafos), e conseqüentemente para a distinção entre estratégias de fluência comunicativa e estilos de fala, bem como para o processamento automático da estruturação multilinear da fala podem ser recuperados automaticamente, não necessitando de ser marcados numa transcrição manual ToBI. Entre esses tipos de informação destacam-se declives de f_0 e de energia intra e inter constituintes entoacionais, reinicializações de f_0 e de energia pós-fronteira, alongamentos finais, mudanças na qualidade da voz e pausas (sobre a importância destes parâmetros numa comparação inter-línguas, veja-se Vaissière 1983). É a partir do contraste estabelecido pela variação contínua destes parâmetros, tanto a nível local como global, que se deteta uma estruturação prosódica multilinear.

3. Variação prosódica

Sabe-se que, tal como a variação de natureza segmental (###3. Variação sociolinguística), lexical e sintática, também a variação prosódica (melódica e temporal) concorre para a distinção entre grupos de falantes do português e entre situações de comunicação oral. No primeiro caso, esta distinção é habitualmente associada a fatores geolinguísticos e sociolinguísticos, nomeadamente: espaço geográfico de origem, género, idade e estatuto sócio-profissional dos falantes. No segundo, a fatores estilísticos dependentes das circunstâncias da situação de produção que, em geral, remetem para os extremos do contínuo de variação entre fala espontânea e fala preparada.

3.1 Variação geolinguística

Relativamente à variação geolinguística da prosódia, no interior da variedade europeia do português, bem como entre variedades nacionais do português, em particular entre português europeu e português brasileiro, são relativamente recentes os estudos comparativos, tanto os de natureza fonética (cf. projeto Atlas Multimédia Prosódico do Espaço Românico – Língua Portuguesa (AMPER-Por)⁵, Moutinho coord., 2000-), como os

⁵ <http://pfonetica.web.ua.pt/AMPER-POR.htm>

desenvolvidos dentro de uma perspectiva fonológica métrica autossegmental (cf. projeto Atlas Interativo da Prosódia do Português (InAPoP)⁶, Frota coord., 2012-2015). Os estudos publicados incidem sobretudo sobre materiais recolhidos em situação experimental para eliciação de frases, e em situação de leitura de frases, e concentram-se na análise de alguns dos tipos de estruturas frásicas mais estudados no português padrão, embora contemplem múltiplos parâmetros de variação prosódica, quer do ponto de vista da produção quer do da perceção: padrões de f_0 , duração e energia, no caso dos trabalhos no âmbito do AMPER-Por (cf., por exemplo, Moutinho/Coimbra 2014; 2015); aspetos rítmicos, fraseamento e extensão dos constituintes entoacionais, tipo e frequência dos contornos nucleares, distribuição de acentos tonais e realização fonética destes (alinhamento com a cadeia segmental), no caso dos trabalhos no âmbito do InAPoP (cf., por exemplo, Frota et al. 2015, e referências aí citadas; Cruz 2013; Vigário/Frota 2003). Progressivamente, estes estudos têm vindo a incluir variedades setentrionais, centro-meridionais e/ou insulares do português europeu, bem como variedades do português brasileiro faladas, principalmente, na costa atlântica. Quanto à caracterização prosódica das variedades africanas do português, não existem (quase) estudos publicados (cf. Santos/Fernandes-Svartman 2014, sobre declarativas (lidas) no português da Guiné-Bissau). Neste capítulo, por limite de espaço, concentrar-nos-emos na variação prosódica no interior do português europeu.

Na maioria dos tipos de estruturas que foram comparados entre variedades do português europeu, encontram-se padrões entoacionais idênticos para funções linguísticas idênticas (para declarativas neutras, cf. Moutinho/Coimbra 2014 e referências aí citadas; para declarativas neutras e focalizadas, imperativas com expressão de ordem e de pedido, interrogativas Qu- e vocativos com expressão de chamamento, cf. Frota et al. 2015). Por exemplo, as declarativas caracterizam-se por um padrão entoacional descendente independentemente das variedades estudadas. Este padrão é formado por um tom alto, H, associado à primeira sílaba acentuada e por um tom baixo, L, associado à última sílaba acentuada da frase, embora possa haver variação no alinhamento do alvo tonal L* entre variedades (mais tardio na variedade de Braga, segundo Vigário/Frota 2003), e na frequência de uso de H+L* L% (dominante nas variedades de Lisboa e do Algarve) e L* L% (mais frequente nas variedades do Porto e do Alentejo, segundo Frota et al. 2015). Esta variação entoacional do português europeu é assim compatível com a boa compreensão da entoação que se sabe existir inter-falantes no interior da variedade europeia do português, tal como é geralmente reconhecido.

As principais diferenças inter-variedades do português europeu são observadas nas interrogativas de *sim-não*: H+L* LH% é usado predominantemente pelos falantes da variedade de Lisboa, L*+H H% por falantes das outras variedades (Frota et al. 2015). Ilustramos para a variedade do Alentejo, na figura 16, os contornos ascendentes L* H% (cf. Cruz 2013) e L*+H H%.

⁶ <http://labfon.letras.ulisboa.pt/InAPoP/presentation.html>

Figura 16. (a) L* H% e (b) L*+H H% em interrogativas de *sim-não* na variedade do Alentejo

Assim, os resultados destes estudos permitem afirmar que as diferenças prosódicas entre variedades do português não se limitam apenas a questões de realização fonética, sem efeitos significativos na interpretação dos contornos nucleares. Podem incluir contrastes de natureza funcional em alguns tipos de frases, i.e., diferentes padrões podem ter um uso idêntico. Para além disso, um mesmo padrão pode ter usos diferentes em algumas variedades do português europeu (cf. Cruz 2013, sobre a não distinção de perguntas neutras e focalizadas na variedade do Algarve, realizadas com o mesmo padrão L*+H H%). São estas as diferenças que podem concorrer para dificuldades de compreensão inter-falantes. No caso das interrogativas de *sim-não*, sabe-se que isso pode acontecer com alguma frequência já que quer em noticiários televisivos quer em fala espontânea as interrogativas de *sim-não* ocorrem muitas vezes com contornos nucleares descendentes (veja-se um exemplo disso na figura 17), tanto na variedade padrão falada em Lisboa (Mata 1990; Moniz et al. 2011), como em outras variedades, nomeadamente na ilha de São Miguel, nos Açores (cf. Moutinho/Coimbra 2015). Nestes casos a distinção prosódica pode ser associada à gama de valores de f_0 , superior nas interrogativas comparativamente às declarativas, e não a contornos nucleares distintivos.

Figura 17. H+L* L% em interrogativa de *sim-não* na variedade de Lisboa (ALERT)⁷

Embora com resultados variáveis entre tipos de frase, os estudos no âmbito do InAPoP (Frota 2012-2015) sugerem ainda que a frequência de distribuição de acentos tonais por palavra prosódica em posição pré-nuclear é menor em Lisboa que nas restantes variedades estudadas, tanto nas meridionais (Alentejo, Algarve), como nas setentrionais (Porto, Braga), que em geral são consideradas mais conservadoras que as restantes. Poder-se-á dizer que a maior ou menor ocorrência de configurações ascendentes e o maior ou menor número de acentos tonais em alguns tipos de frase parecem convergir para a sensação da diversidade sonora do país: nem todas as variedades do português europeu terão uma pronúncia seca e um tom cerrado, como habitualmente parece aos ouvidos de falantes de variedades africanas e brasileiras do português, sobretudo no caso da variedade padrão lisboeta.

⁷ O corpus ALERT é constituído por telejornais da RTP (Neto et al. 2003; Meinedo/Neto 2003).

3.2 Variação social

Comparativamente à variação inter-falantes no espaço geográfico, que apresenta um número crescente de publicações, a área da variação social da prosódia tem vindo a receber menos atenção, em português, tal como em muitas outras línguas. É principalmente no âmbito da investigação sobre fala espontânea, e em particular sobre estilos de fala, exclusivamente na variedade padrão falada em Lisboa, que têm vindo a ser sugeridas pistas para uma distinção prosódica inter-falantes, por influência de variáveis como género, grupo etário e/ou estatuto sócio-profissional dos falantes.

Algumas das características da variação por género observadas na fala espontânea de adultos são a seguir indicadas. Há, em média, sequências inter-pausas menos longas e menos tempo de pausa silenciosa (menos pausas, mais pausas breves) no discurso feminino em situação de conversa (sem alternância de tomadas de vez). Há, em média, mais tempo de pausa preenchida no discurso masculino (pausas preenchidas em maior número) na mesma situação. As diferenças apontadas são associadas a uma maior rapidez do discurso oral feminino (Freitas 1990). O tom fronteira LH% é em geral mais frequente para as falantes do género feminino em situação de relato espontâneo. Os acentos bitonais tendem a ser mais usados pelo género feminino (e.g., H+L*, L+H*) e L* pelo masculino (Mata/Moniz/Batista no prelo), o que é consistente com observações feitas para outras línguas (cf. Clopper/Smiljanic 2011, sobre a maior frequência de tons complexos no discurso feminino).

Também na fala espontânea dos adolescentes (14-15 anos) se observa variação prosódica em função da variável género. Os rapazes distinguem-se das raparigas por uma velocidade de elocução e de articulação⁸ significativamente mais baixa (em média, há respetivamente 4,9/8,1 sílabas por segundo para os rapazes, em contraste com 6,6/9,6 para as raparigas, em constituintes similares a frase sem disfluências). Estas diferenças entre géneros podem ser interpretadas em linha com as diferenças apontadas por Freitas (1990) para adultos. O contraste entre constituintes entoacionais menores e maiores é em geral mais forte para os rapazes do que para as raparigas: estas produzem maior percentagem de ips que os rapazes; estes produzem maior percentagem de IPs que as raparigas (Mata/Moniz/Batista no prelo). Os padrões finais descendentes-ascendentes (de tipo H+L* LH%) são mais raros para os rapazes que para as raparigas, e o inverso se passa com os padrões ascendentes-descendentes (Mata 1999), o que é consistente com o observado acima para os adultos.

Relativamente à variação inter-falantes por grupo etário/estatuto, até agora apenas as diferenças entre adultos (professores) e adolescentes (alunos de 14-15 anos) foram abordadas numa perspetiva comparativa. As principais diferenças entre adultos e adolescentes em situação de relato espontâneo incluem (Mata et al. 2014b; Mata/Moniz/Batista 2014; no prelo): a velocidade de elocução e de articulação, significativamente mais baixa para os adultos (em média, há respetivamente 4,9/6,7 sílabas por segundo para os adultos vs. 5,8/8,8 para os adolescentes, em constituintes

⁸ Ambas são calculadas em sílabas por segundo: a velocidade de elocução corresponde ao número de sílabas dividido pela duração total de fala (incluindo silêncios), e a de articulação, ao número de sílabas dividido pela duração de fala, excluindo silêncios.

similares a frase sem disfluências), e o rácio de fonação⁹, em geral significativamente mais elevado para os adultos (72,3% vs. 65% para os adolescentes). Incluem também o uso de contornos nucleares de tipo alto/ascendente associados a valores pragmáticos de continuação (principalmente (L+)H* H) em enunciados declarativos, significativamente mais frequente para os adultos, e a frequência de processos fonético-fonológicos (principalmente, desvozeamento/elisão de material pós-tónico, cf. figuras 18 e 19; efeitos de laringalização¹⁰, cf. figura 20) pré-fronteira de constituinte entoacional, mais elevada para os adolescentes. Estes processos pré-fronteira são difíceis de processar automaticamente e, com frequência, originam segmentações automáticas erróneas, sobretudo para a fala dos adolescentes (Mata et al. 2014b).

Figura 18. Elisão de vogal pós-tónica em *segundo* (CPE-FACES)

Figura 19. Elisão de vogal postónica e fricativação da última oclusiva em *temp(o)* (CPE-FACES)

Figura 20. Efeitos de laringalização final em *escola* (CPE-FACES)

Quanto à variação em função do estatuto dos falantes, destaca-se o estudo comparativo de Delgado-Martins e Freitas (1991) sobre índices de marcação temporal em adultas de diferentes grupos sócio-profissionais (professoras universitárias e pivôs de telejornais). Este estudo demonstra que a produção de pausas silenciosas e preenchidas permite distinguir as estratégias usadas por professoras e jornalistas na fala espontânea: as primeiras produzem mais pausas silenciosas (21% vs. 13%), mas muito menos pausas preenchidas do que as segundas (6% vs. 32%). As autoras apontam no sentido de as professoras usarem mais pausas silenciosas com o intuito de clarificar a mensagem e de as jornalistas hesitarem mais no planeamento verbal.

⁹ O rácio de fonação corresponde a 100% a multiplicar pela duração de fala sem silêncios, dividido pela duração de fala incluindo silêncios (Cucchiari/Strik/Boves 2002).

¹⁰ O termo ‘laringalização’ corresponde à tradução do inglês ‘creaky voice’ ou ‘laryngealized voice’ (Andrade 1990).

Em conjunto, estes dados sobre a variação de estratégias interfalantes numa mesma situação de comunicação demonstram que a variação prosódica é uma dimensão importante para caracterizar o que é comum à variedade europeia do português e contribui para a codificação da identidade dos falantes de acordo com grupos sociais de pertença (incluindo, género, grupo etário, grupo sócio-profissional e variedade regional).

3.3 Variação estilística

Os dados acumulados nos últimos anos sobre a variação estilística da prosódia no português europeu demonstram que a variação entoacional e temporal permite distinguir entre diferentes estilos de fala, configura diferentes estratégias de fluência, e põe em evidência diferenças entre géneros e grupos etários nas estratégias de adaptação dos falantes a atividades de produção oral específicas. É à comparação das principais características da variação prosódica entre estilos de fala (intra e interfalantes) que se dedica o resto desta secção.

O contínuo de variação estilística fala espontânea-fala preparada (designações que remetem para os seus dois extremos) tem sido central na investigação sobre variação prosódica em português europeu, como em muitas outras línguas. Tradicionalmente, a investigação sobre estilos de fala concentra-se nas diferenças entre atividades de produção oral espontânea e de leitura de texto (para uma visão geral e uma discussão, cf. Eskénazi 1993; Barry 1995; Hirschberg 2000). Em português europeu, para além da leitura de texto (Freitas 1990), outros tipos de produção oral com preparação prévia, mas sem recurso à leitura de materiais escritos, têm sido estudados numa perspectiva comparativa. Por exemplo, relativamente à extensão de constituintes similares a frase (do inglês *sentence-like units*, i.e., constituintes delimitados por sinais de pontuação e, na sua maioria, por fronteiras tonais), sabe-se que os tipos de produção que se caracterizam por uma menor extensão do planeamento e uma natureza mais interativa se distinguem tipicamente por terem constituintes similares a frase de menor extensão (e.g. diálogos espontâneos), bem como velocidades de elocução e de articulação mais baixas e um rácio de fonação mais elevado. Com resultados variáveis, o inverso se passa nos tipos de produção que divergem no grau de espontaneidade e de planeamento prévio e têm uma natureza mais expositiva (e.g. aulas universitárias). Comparativamente a um noticiário televisivo, por exemplo, a extensão dos constituintes é menos acentuada numa aula universitária, sem dúvida como reflexo da atenção dada pelo falante à compreensão de conceitos, com paráfrases, explicação e exemplificação detalhadas, em suma, ao tempo que é investido na construção do conhecimento dos participantes. A extensão dos constituintes tende a aumentar como efeito do esforço do falante para apresentar grande quantidade de informação, de forma clara e bem estruturada, num curto intervalo de tempo, muitas vezes com recurso à leitura de materiais escritos (e.g., em noticiário televisivos ou em discursos no parlamento europeu), o que contrasta com o que é habitual numa aula universitária (cf. Moniz 2013, para uma comparação entre diferentes produções orais). Tenderá a diminuir como efeito de fatores de orientação para o ouvinte (e.g. em diálogos dirigidos a crianças). Veja-se o quadro 1.

| Corpora | Média de palavras |
|--|-------------------|
| Diálogos adulto-criança | 3 |
| Diálogos em formato <i>map-task</i> | 6 |
| Relato espontâneo de professores de 3º Ciclo | 8 |
| Aula de 3º Ciclo | 17 |
| Aulas universitárias | 18 |
| Noticiários televisivos | 22 |
| Sessões do parlamento europeu | 29 |

Quadro 1. Média de palavras por constituinte similar a frase em diferentes corpora

Também a extensão do fraseamento prosódico em relatos espontâneos e exposições escolares (previamente preparadas, mas sem recurso à leitura de materiais escritos), de adultos e adolescentes, surge como um indício do grau de planeamento entre os estilos de fala: há na fala preparada uma divisão em constituintes de maior extensão e regularidade (cf. linhas 3 e 4 do quadro 1). Surge ainda como marca de uma menor unidade da fala preparada dos alunos: apesar da variação inter-falantes, há na fala preparada do professor maior regularidade relativamente à extensão do fraseamento prosódico. Este tende a integrar mais palavras por constituinte do que os alunos, traçando assim blocos mais amplos e mais fortemente marcados do que na fala espontânea. Pode até especular-se sobre o facto de o discurso preparado libertar parcialmente o falante no que diz respeito ao conteúdo, favorecendo uma maior concentração no controlo da forma entoacional (cf. Mata 1999).

No resto desta secção, incluem-se na variação entre estilos de fala as diferenças entre diálogos espontâneos e aulas universitárias (Trancoso et al. 2008; Moniz et al. 2011; Moniz 2013; Moniz et al. 2014b), e as diferenças entre relatos espontâneos e apresentações orais escolares (Mata 1999; Moniz 2006; Moniz et al. 2014a; Mata et al. 2014b; Mata/Moniz/Batista 2014; no prelo). Tal como referido acima, estas diferenças estão relacionadas com a quantidade de planeamento prévio envolvido nos tipos de produção, em função da clareza da informação que se pretende transmitir e do grau de interação entre falante e ouvinte(s). Concentramo-nos na forma como se manifestam nos padrões de organização prosódica, em termos de: (a) distribuição de eventos tonais (acentos nucleares e fronteiras tonais); (b) variação da gama de valores de f_0 e de energia; (c) distribuição de pausas silenciosas; (d) distribuição de eventos disfluentes e padrões de fluência.

(a) Distribuição de eventos tonais (acentos nucleares e fronteiras tonais)¹¹

A frequência de uso de padrões entoacionais descendentes vs. ascendentes em enunciados declarativos varia entre estilos de fala, com diferenças entre grupos etários. É disso exemplo no português europeu a variação entre relatos espontâneos e exposições preparadas de adultos e adolescentes.

¹¹ Está ainda por concluir uma análise deste tipo de variação em diálogos espontâneos e aulas universitárias. Por essa razão, apresentamos nesta secção as principais diferenças entre relatos espontâneos e exposições preparadas.

Nos adultos, há mais acentos nucleares de tipo baixo/descendente – principalmente (H+)L* – e fronteiras L% na exposição preparada; os acentos nucleares de tipo alto/ascendente – principalmente (L+)H* – e as fronteiras H% são mais frequentes e usados em proporções mais equilibradas no relato espontâneo. Esta distribuição das declarativas ascendentes é convergente com o que tem sido observado para outras línguas em monólogos espontâneos e narrativas espontâneas (Blaauw 1995; Wennerstrom 2001), e interpretado como efeito quer do uso de estratégias para assinalar a não-terminalidade dos enunciados (em geral correlacionadas com tons finais H e consideradas mais frequentes na fala espontânea) quer do uso de estratégias para envolver os ouvintes, apelando e mantendo a atenção destes (consideradas mais frequentes em atividades de fala de tipo narrativo).

Nos adolescentes, passa-se o inverso: os contornos nucleares de tipo baixo/descendente predominam nos enunciados declarativos do relato espontâneo; os contornos nucleares de tipo alto/ascendente aumentam nos enunciados declarativos da exposição preparada. Esta variação é independente do género dos falantes. Por conseguinte, não se confirma no português europeu, a afirmação corrente para outras línguas de que há mais declarativas ascendentes na fala espontânea dos adolescentes, em particular nas raparigas. A maior frequência de padrões ascendentes na fala espontânea dos adultos é um dos fatores que os distingue dos adolescentes.

Refira-se a propósito que as estratégias de distribuição de tons ascendentes *vs.* descendentes entre situações comunicativas estão correlacionadas com a frequência de efeitos fonético-fonológicos comuns no português europeu em posição pré-fonteira, nomeadamente: inserção de [@], correlacionada com fronteiras H (cf. figura 21), e efeitos de aspiração final (cf. figura 22) são muito mais frequentes nas exposições preparadas; desvozeamento/apagamento de vogal postónica (cf. figuras 18 e 19 acima) e efeitos de laringalização final, correlacionados com fronteiras L (cf. figura 23), são mais comuns nos relatos espontâneos.

Figura 21. Inserção final de [a] em *amor* (CPE-FACES) Figura 22. Efeitos de aspiração final em *consistia* (CPE-FACES)

Figura 23. Efeitos de laringalização final em *vida* (CPE-FACES)

(b) Variação da gama de valores de f_0 e de energia

Também a variação da gama de valores de f_0 em que se inscrevem os contornos entoacionais e a variação dos valores de energia são condicionadas pelas circunstâncias das situações de comunicação. Declives

de f_0 , valores de f_0 máxima, f_0 mínima e desvios padrão, bem como declives de energia encontram-se entre os parâmetros que mais concorrem para esse efeito. Os diálogos apresentam valores mais elevados de energia; as aulas universitárias valores mais elevados de f_0 . O mesmo tipo de contraste se observa entre relatos espontâneos e exposições preparadas de adultos.

Sabe-se igualmente que os falantes podem aplicar diferentes estratégias no uso destes parâmetros em diferentes estilos de fala, tanto em contextos fluentes como disfluentes, o que se associa a uma maior expressividade e carisma dos falantes. Por exemplo, nas aulas universitárias, o falante avaliado como mais expressivo (S6 no Quadro 2) é o que conjuga com mestria declives de f_0 e de energia mais elevados na reposição da fluência e maior controlo de diversas medidas temporais (velocidades de elocução e de articulação mais elevadas) que contribuem para a dinâmica da interação com os alunos, correlatos que têm sido associados a estratégias de fluência e de competência linguística dos falantes.

| Parâmetros | Média para o falante S6 | Média para os restantes falantes masculinos |
|---|-------------------------|---|
| \neq de f_0 da disfluência para a palavra seguinte (ST) | 0,81 | 0,74 |
| Declive de energia da palavra seguinte à disfluência | 1,35 | 0,77 |
| Média de energia em constituintes similares a frase (Hz) | 54,2 | 44,2 |

Quadro 2. Principais diferenças de f_0 e de energia por constituinte similar a frase em aulas universitárias (LECTRA)¹²

(c) Distribuição de pausas silenciosas

Na leitura, as pausas silenciosas ocorrem sobretudo em posições sintáticas hierarquicamente superiores, ou seja, em fronteiras de frase e oração; na fala espontânea, sobretudo entre e/ou dentro de constituintes sintáticos (Freitas 1990). A distribuição de pausas silenciosas não permite apenas distinguir leitura de fala espontânea. Nas aulas universitárias, há mais pausas silenciosas do que nos diálogos, o que se correlaciona, por um lado, com a dinâmica de um diálogo e, por outro, com a multifuncionalidade das pausas silenciosas nas aulas, e.g., dar a palavra, enfatizar informação, levar os alunos a refletirem sobre o tópico em análise antes da introdução de um novo tópico, mudar de tópico. Por outro lado ainda, a presença/ausência de interlocutores também afeta a distribuição de pausas silenciosas: numa aula via internet (cf. falante S5 na figura 24) a ocorrência de pausas silenciosas tende a ser reduzida; o inverso acontece se o professor estruturar a sua aula em interação constante com os alunos, numa dinâmica similar à de um diálogo (cf. falante S6 na figura 24).

¹² O corpus LECTRA é constituído por aulas universitárias (Trancoso et al. 2008; ISLRN | 298-379-572-530-5).

Figura 24. Percentagem de tempo de silêncio e de fala de dois professores universitários (S5 = aula online; S6 = aula presencial) (LECTRA)

O mesmo tipo de contraste se verifica quando se comparam exposições preparadas e relatos espontâneos produzidos quer por adultos quer pela maioria dos adolescentes. A maior frequência de pausas silenciosas nas exposições preparadas, em conjunto com a variação de outros parâmetros melódicos e temporais, pode ser interpretada como mais um indicador das diferenças de inteligibilidade entre estilos de fala.

(d) Distribuição de eventos disfluentes e padrões de fluência

Sabe-se para o português europeu (Moniz 2006; 2013) que os fenómenos ditos disfluentes, como pausas preenchidas, alongamentos, repetições, fragmentos, substituições, apagamentos, marcadores de edição e sequências complexas (cf. exemplos i-viii, a seguir), consoante as propriedades prosódicas que exibem, podem ser percetivamente julgados como fluentes ou disfluentes.

- i. <aa> *Houve um marco fundamental na evolução da indústria informática* (pausa preenchida, sem conteúdo lexical, essencialmente uma vogal central alongada)
- ii. *regularização de existências e= o que é que aconteceu agora?* (alongamento de material linguístico)
- iii. *é preciso codificar <não é> não é direto* (repetição de material linguístico)
- iv. *porque não há <direç-> direcionalidade na relação* (fragmento ou truncção de um item lexical)
- v. *que, aliás, <sai> saiu na vossa ficha* (substituição de material linguístico)
- vi. *<significa isto que se nós considerarmos nós temos aqui> Ah, e no fim, e no fim, diz aí que vocês tinham ainda um stock de cento e cinquenta traves* (apagamento ou abandono de material linguístico e início de uma nova ideia)
- vii. *esta cobra <que une que vai que não é que une> que vai depois* (marcador de edição ou expressão de reformulação explícita)
- viii. *O ano passado houve uns colegas vossos da matemática que <fizeram o projeto só qua-> queriam fazer o projeto quase só com strings.* (sequência complexa, i. e., combinatórias de mais do que um tipo de disfluência)

Numa tarefa de classificação de disfluências como mecanismos de fluência/disfluência (com recurso a Árvores de Regressão e Classificação), as propriedades prosódicas são as pistas linguísticas mais salientes – nomeadamente, o fraseamento prosódico e os contornos de f_0 . 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 (cf. figura 25), por oposição aos que são produzidos no interior de um constituinte ou formam por eles próprios um

constituente, exibindo contornos descendentes ou efeitos de glotalização (cf. figura 26). Poder-se-á reafirmar, tal como para outras línguas, que os falantes controlam diferentes aspetos 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.

Figura 25. Exemplo classificado percetivamente como fluente (CPE-FACES)

Figura 26. Exemplo classificado percetivamente como disfluente (CPE-FACES)

A comparação intra-corpora (relatos espontâneos *vs.* exposições preparadas não-lidas) aponta para uma maior frequência de eventos disfluente na fala espontânea dos adultos, em contraste com uma maior frequência de eventos disfluente na fala preparada dos adolescentes, o que pode ser associado a graus de adaptação a uma tarefa especificamente escolar. Como na exposição preparada o esforço de monitorização e planeamento por parte dos falantes é significativamente maior, é nesta situação que são produzidas mais disfluências. Os professores, porém, demonstram ter um comportamento diferente, produzindo mais pausas preenchidas na situação de relato espontâneo e com uma percentagem de tempo disfluente superior nesta situação comunicativa, o que sugere que estes falantes evidenciam maior habituação e treino de exposição de conteúdos do que propriamente no relato de experiência pessoais.

A comparação inter-corpora aponta efeitos de estilo nos padrões temporais e mesmo na marcação prosódica por contraste através de um aumento dos valores de energia e de f_0 da disfluência para a reposição da fluência na palavra seguinte, entre aulas universitárias e diálogos. Embora as pausas preenchidas sejam a categoria mais representativa em ambos os corpora, as restantes categorias apresentam uma distribuição distinta. Nas aulas, as sequências complexas (e.g., repetições e substituições em conjunto, utilizadas para procura/precisão lexical) são mais frequentes do que as repetições por si só, enquanto nos diálogos ambas têm distribuições similares. Nos diálogos, os fragmentos correspondem a mais do dobro dos fragmentos produzidos nas aulas e os apagamentos são residuais. Estas diferenças na distribuição das categorias disfluente podem ser interpretadas em função da natureza dos diálogos em análise, nomeadamente das restrições temporais a que os falantes estão sujeitos em diálogos para

indicação de percursos em mapas, com recurso mais frequente a categorias como repetições e fragmentos e menos a sequências complexas e apagamentos. Os padrões temporais também apontam para a natureza mais dinâmica dos diálogos por comparação com as aulas, com produção de menos palavras, tanto em frases fluentes como em frases que contêm disfluências. O encadeamento das interações comunicativas num diálogo está sujeito a restrições temporais, evidentes também na duração dos silêncios, na disfluência e nos próprios contextos adjacentes. Uma vez mais, todas as unidades referidas são mais breves nos diálogos do que nas aulas. Tanto os diálogos como as aulas universitárias exibem aumentos de f_0 e de energia (marcação prosódica por contraste) entre a disfluência e a reposição da fluência para a maioria das categorias disfluentes, embora com diferentes graus de contraste. Porém, mesmo a referida estratégia de marcação prosódica por contraste da disfluência para a reposição da fluência está sujeita a variação inter-corpora, sendo esta marcação mais forte nas aulas que nos diálogos. Nas aulas, pistas de f_0 e de energia são produzidas por todos os falantes, para a maioria das categorias, tanto para as disfluências como para os contextos adjacentes. O conjunto de padrões apresentado é um contributo para a diferenciação entre estilos de fala, nomeadamente entre fala espontânea e fala preparada não-lida.

4. Prosódia e processamento automático

No que respeita à interface com o processamento automático de fala, uma das áreas mais produtivas nos últimos anos, a transversalidade da análise prosódica tem tido efeito no renovar de metodologias diversas de processamento da informação prosódica, contribuindo, por um lado, para uma sistematização do ponto de vista linguístico e, por outro, para o melhoramento dos sistemas de síntese e de reconhecimento de fala.

O enriquecimento automático de transcrições de fala com um conjunto de eventos designados no inglês *structural metadata events* (Liu et al. 2006; Ostendorf et al. 2008) – i.e., sinais de pontuação e maiúsculas em fronteira de frase, bem como disfluências e outros marcadores discursivos – contribui significativamente para a legibilidade de uma sequência de palavras obtida através de um reconhecedor automático de fala. Este processo é crucial para diversas aplicações (e.g., legendagem automática de noticiários televisivos, estruturação de tópicos e sumário de notícias para pesquisa, exercícios para a aprendizagem de L1/L2 mediada por computador) e para a própria arquitetura do reconhecedor. Frequentemente, o reconhecedor de fala está integrado numa sequência de módulos que incluem segmentar áudio, grafar a maiúsculas e minúsculas, identificar zonas de disfluência, sumariar áudio. Por conseguinte, o enriquecimento de transcrições de fala pode ser entendido como uma tarefa de estruturação de uma sequência de palavras em diferentes unidades linguísticas, ou seja, como uma estruturação multilinear, abrangendo módulos da gramática distintos.

Para este processo tem sido utilizado um conjunto alargado de pistas, reflexo da estruturação multilinear acima referida. Estas pistas estão muito para além das lexicais (n-gramas de palavras, ou seja uma ou mais palavras) extraídas da saída do reconhecedor, ou das acústicas extraídas do módulo de processamento de áudio (identificação de segmentos correspondentes a fala

vs. outros tipos de segmentos nos quais se incluem silêncios, música, *inter alia*; identificação de falantes, etc.). No trabalho desenvolvido para o português, à semelhança do que acontece para outras línguas, tem sido dado especial destaque às pistas prosódicas.

O enriquecimento de transcrições automáticas depende de: i) transcrições manuais, ii) transcrições produzidas pelo reconhecedor automático e iii) análise do sinal acústico (para mais informações sobre este processo, veja-se Batista et al. 2012a e 2012b; Moniz 2013; Moniz et al. 2014b; Cabarrão et al. 2015). A anotação manual é complexa sobretudo no caso da anotação de corpora de fala espontânea, uma vez que implica pontuar fala, por um lado, e identificar disfluências e marcadores discursivos, por outro – tarefas que, como se sabe, afetam a concordância inter-anotadores em distintos corpora do português europeu (cf. Batista 2011 e Cabarrão et al. 2014 para um corpus de notícias televisivas; Moniz 2013 para um de aulas universitárias). A saída do reconhecedor já identifica e delimita temporalmente fones, sílabas, palavras e silêncios. Com base nesta e na informação contida nas transcrições manuais, as sequências são organizadas numa estruturação hierárquica que contempla ainda unidades similares a frases e atos de fala. A partir desta estruturação, são extraídas pistas acústicas segmentais e suprasegmentais, a saber, valores mínimos, máximos, médias, medianas, desvios padrão e declives de f_0 e de energia, bem como a duração de cada uma das unidades. Os valores absolutos dos declives de f_0 e de energia são convertidos em 3 categorias básicas (ascendente, descendente e *plateau*) e em 6 combinatórias (e.g., ascendente-descendente, ascendente-*plateau*). Rácios de duração de unidades e de silêncios adjacentes são também calculados com base em três valores: maior, igual a, menor.

As experiências já realizadas para o português europeu demonstram que sinais de pontuação, tipos de disfluência, reposição de fluência e marcadores discursivos, em geral, podem ser preditos e caracterizados com recurso apenas à informação prosódica.

Quanto aos sinais de pontuação, na predição de pontos finais destacam-se as seguintes pistas: i) contorno descendente na palavra antes de um ponto final; ii) nível estacionário de energia na mesma palavra; iii) duração relativa entre essa palavra e a seguinte; e iv) grau superior de confiança em relação à palavra seguinte. Este conjunto de pistas é ilustrativo do comportamento de uma declarativa neutra no português europeu (cf. secção 2). Os pontos de interrogação, por sua vez, são caracterizados por dois padrões diferenciados: i) contorno de f_0 ascendente na palavra antes de um ponto de interrogação e declive de energia ascendente nessa e na palavra seguinte; ii) contorno de f_0 estacionário na palavra antes de um ponto de interrogação e declive de energia descendente nessa mesma palavra. Os pontos de interrogação surgem, assim, como uma classe difícil de detetar (e.g., Moniz et al. 2011; Margolis 2011), por poderem corresponder a distintos padrões prosódicos em função dos tipos de interrogativas, como se viu na secção 2. Quanto às vírgulas, nas experiências até agora realizadas para o português europeu, tal como para outras línguas, elas são o evento mais difícil de predizer prosodicamente. Tal como expetável, são sobretudo classificadas com base em pistas morfo-sintáticas, não sendo claramente desambiguadas por meio de pistas prosódicas.

As pistas que mais contribuem para a predição da reposição de fluência a seguir a uma sequência disfluente integram: i) duas palavras contíguas idênticas (única informação lexical, passível de ser identificada acusticamente); ii) subida dos níveis de f_0 e de energia na palavra que inicia uma reposição de fluência e um contorno estacionário na palavra anterior; iii) grau de confiança dado pelo reconhecedor sobre a palavra que inicia a reposição superior ao da disfluência propriamente dita.

Experiências mais recentes (Cabarrão et al. 2015) sobre disfluência e marcadores discursivos em início de enunciado mostram que aquelas são mais bem preditas do que estes. Para isso concorre a diversidade de padrões prosódicos associados a tipos específicos de marcadores discursivos, nomeadamente, desacentuação vs. acentuação, com variação de nível e gama de f_0 (tons H escalonados acima de outros tons H subsequentes; contornos *plateau* em espaço tonal intermédio relativamente aos constituintes subsequentes). Por exemplo, o marcador conversacional “agora” é habitualmente associado a um acento de tipo H+L e fraseado separadamente, formando por si só um constituinte entoacional, provavelmente para indicar uma mudança do tópico discursivo, por oposição a “portanto”, que é geralmente desacentuado. Está em curso uma análise mais aprofundada da relação dos padrões prosódicos com a estrutura do discurso e o estatuto informacional destes marcadores.

Para além do enriquecimento de transcrições automáticas, outras linhas de investigação têm vindo a ser desenvolvidas: a predição de proeminência prosódica e a deteção de emoções. A primeira com o objetivo de adicionar ao módulo prosódico informação relativa a proeminência prosódica (acentuação/desacentuação) com base no sistema ToBI. A segunda para um melhor conhecimento dos eventos paralinguísticos de modo a aferir o conteúdo emocional das interações comunicativas e a sua variação social.

Para treinar um modelo de proeminência prosódica do português, foi usado um corpus de fala espontânea e preparada (não lida) de adultos e adolescentes. Adicionalmente, foram utilizados modelos do inglês americano para aferir da validade de identificação de proeminência prosódica entre línguas. Os modelos do inglês permitem prever proeminência prosódica com 71% de acerto nos dados do português e os modelos treinados especificamente com dados do português uma taxa de 77% de acerto, o que representa uma subida de 6% absolutos. Estes resultados demonstram que os eventos proeminentes são detetados com uma percentagem considerável de acerto de uma língua para a outra, como se exemplifica na figura 30. Demonstram ainda um efeito de idade/estatuto dos falantes, uma vez que a predição de proeminência é sempre melhor nos adultos do que nos adolescentes. Esta linha de investigação sobre predição de proeminência pode, assim, contribuir para a discussão de correlatos acústico-prosódicos universais vs. específicos de uma língua ou para avaliar a variação dependente do domínio de fala e de variáveis sócio-linguísticas.

Figura 27. Exemplo de detecção automática de proeminência com o AuToBI. As primeiras 3 frases correspondem à transcrição manual; as 3 seguintes à aplicação do modelo do inglês ao português; as 3 últimas à aplicação do modelo treinado especificamente para o português.

No âmbito da detecção de emoções para o português europeu, o stress é o que tem merecido maior atenção, uma vez que o medo de falar em público (glossofobia) é uma das fobias mais comuns nos adultos (Miller/Stone 2009). O principal objetivo é o desenvolvimento de aplicações para o treino de fala em público, com monitorização dos níveis de stress.

Para a predição de níveis de stress foi usado um corpus de fala com ocorrência natural, constituído por apresentações de teses de mestrado em provas públicas (Aguiar et al. 2014). Concorre para essa predição um vasto conjunto de pistas acústico-prosódicas. Entre as pistas prosódicas mais informativas destacam-se as seguintes: parâmetros de qualidade da voz associados a voz laringalizada, nomeadamente *jitter* (perturbações de f_0) e *shimmer* (perturbações de energia) ao longo dos constituintes similares a frase, e movimentos finais de f_0 sistematicamente mais baixos do que o expectável, constituinte após constituinte (cf. Julião et al. 2015). Sabe-se que também em exposições escolares de adolescentes, o stress ou a insegurança se associam a um nível de registo substancialmente mais baixo e a uma redução da gama de f_0 , o que se traduz numa nítida impressão de falta de controlo da situação (Mata 1999).

A detecção de emoções é uma área de especial interesse para interações comunicativas com agentes virtuais ou robôs, na área da robótica social, ou para sistemas de diálogo no domínio da *Intelligent Voice Response* (IVR), ou para a síntese de fala, ou ainda para o desenvolvimento de modelos de compreensão no âmbito do reconhecimento de fala. Os detetores de emoções desenvolvidos para o português europeu têm sido integrados numa plataforma web designada *SPeech Analytics* (SPA), desenvolvida pelo INESC-ID (Batista et al. 2016).

5. Conclusão

Neste capítulo procurámos integrar algumas das formas e funções da variação da entoação no português europeu, considerando ainda as funções de outros recursos prosódicos em diferentes situações de comunicação. Partindo de funções clássicas da variação entoacional (e.g., declarativas vs. interrogativas) que têm sido objeto de maior número de estudos, destacámos os recursos prosódicos que podem desempenhar funções pragmático-discursivas (e.g., em diferentes tipos de perguntas e respostas; mecanismos de disfluência e reposição de fluência; marcadores discursivos; expressão de emoções), estilísticas (fala espontânea vs. preparada e diferenças entre tipos de fala espontânea) e sociolinguísticas (e.g., adultos vs. adolescentes; falantes masculinos vs. femininos) no português europeu, com base em excertos de corpora de natureza mais/menos expositiva ou dialógica.

Para o português europeu, são escassos os estudos de prosódia em contexto de interação. São também escassos os estudos comparativos de prosódica inter-estilos e inter-falantes, em função de fatores como idade e género. Em trabalho futuro, deverá alargar-se a análise da variação

prosódica a mais situações de interação espontânea (humano-humano, humano-agente virtual e humano-robô), comparando diferentes grupos etários (crianças, adolescentes, adultos e idosos) e abrangendo mais fatores de variação (paralinguísticos, sociais e estilísticos). Esta área, embora complexa, é crucial para um conhecimento mais aprofundado das propriedades de estruturação do diálogo e para a integração destas em sistemas de diálogo, por exemplo, contribuindo para modelos de compreensão no processamento automático de fala.

6. Referências

- Aguiar, Ana/Kaiseler, Mariana/Cunha, Mariana/Silva, Jorge/Meinedo, Hugo/Almeida, Pedro (2014), *VOCE Corpus: Ecologically Collected Speech Annotated with Physiological and Psychological Stress Assessments*, in: Nicoletta Calzolari/Khalil Choukri/Thierry Declerck/Hrafn Loftsson/Bente Maegaard/Joseph Mariani/Asuncion Moreno/Jan Odijk/Stelios Piperidis (Edd.) Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, 1568-1574.
- Andrade, Amália (1990), *Aspectos fisiológicos e aerodinâmicos da produção da fala*, in Maria Helena Mira Mateus/Amália Andrade/Maria do Céu Viana/Alina Villalva (edd.), *Fonética, Fonologia e Morfologia do Português*, Lisboa, Universidade Aberta, 57-96.
- Avesani, Cinzia/Hirschberg, Julia/Prieto, Prieto (1995), *The intonational disambiguation of potentially ambiguous utterances in English, Italian and Spanish*, in: Proceedings ICPhS 95, Stockholm, vol. 1, 174-177.
- Batista, Fernando (2011), *Recovering capitalization and punctuation marks in speech transcripts*, Dissertação de Doutorado, Instituto Superior Técnico.
- Batista, Fernando/Moniz, Helena/Trancoso, Isabel/Mamede, Nuno (2012a), *Bilingual Experiments on Automatic Recovery of Capitalization and Punctuation of Automatic Speech Transcripts*, IEEE Transactions on Audio, Speech, and Language Processing, 20 (2), 474-485.
- Batista, Fernando/Moniz, Helena/Trancoso, Isabel/Mamede, Nuno/Mata, Ana Isabel (2012b), *Extending automatic transcripts in a unified data representation towards a prosodic-based metadata annotation and evaluation*, Journal of Speech Sciences, 2, 115-138.
- Batista, Fernando/Curto, Pedro/Trancoso, Isabel/Abad, Alberto/Ferreira, Jaime/Ribeiro, Eugénio/Moniz, Helena/de Matos, David/Riberio Ricardo (2016), *SPA: web-based platform for easy access to speech modules*, in: LREC 2016.
- Barry, William (1995), *Phonetics and phonology of speaking styles*, in Kjell Elenius/Peter Branderud (Edd.), in: *Proceedings of ICPhS 1995*, Stockholm, Sweden, vol.2, 4-10.
- Blaauw, Eleanor (1995), *On the perceptual classification of spontaneous and read speech*, Utrecht, LED.
- Beckman, Mary/Pierrehumbert, Janet (1986), *Intonational structure in Japanese and English*, *Phonology Yearbook 3*, Cambridge, CUP, 255-309.

- Boersma, Paul/Weenink, David (2013), *Praat: doing phonetics by computer* [Computer program]. Version 5.3.56, retrieved 15 September 2013 from <http://www.praat.org/>
- Bruce, Gösta (1977), *Swedish word accents in sentence perspective*, Lund: Gleerup.
- Cabarrão, Vera (2013), *Respostas afirmativas em diálogos espontâneos em português europeu (interface prosódia-sintaxe-discurso)*, Dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Cabarrão, Vera/Moniz, Helena/Batista, Fernando/Ribeiro, Ricardo/Mamede, Nuno/Meinedo, Hugo/Trancoso, Isabel/Mata, Ana Isabel/de Matos, David (2014), *Revising the Annotation of a Broadcast News Corpus: a Linguistic Approach*, in: Nicoletta Calzolari/ Khalil Choukri/Thierry Declerck/Hrafn Loftsson/Bente Maegaard/Joseph Mariani/Asuncion Moreno/Jan Odijk/Stelios Piperidis (Edd.) Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, 3908-3913.
- Cabarrão, Vera/Moniz, Helena/Ferreira, Jaime/Batista, Fernando/Trancoso, Isabel/Mata, Ana Isabel/Curto, Sérgio (2015), *Prosodic classification of discourse markers*, in: Proceedings of ICPhS 2015, Glasgow, Scotland.
- Clopper, Cynthia/Smiljanic, Rajka (2011), *Effects of gender and regional dialect on prosodic patterns in American English*, *Journal of Phonetics*, 39, 237–245.
- Cruz, Marisa (2013), *Prosodic variation in EP: phrasing, intonation and rhythm in Central-Southern varieties*, Dissertação de Doutoramento, Universidade de Lisboa.
- Cucchiari, Catia/Strik, Helmer/Boves, Lou (2002), *Quantitative assessment of second language learner's fluency: comparisons between read and spontaneous speech*, *Journal of the Acoustic Society of America*, 111, 2862–2873.
- Delgado-Martins, Maria Raquel/Freitas, Maria João (1991), *Temporal structures of speech: reading news on TV*, in: Proceedings of the ETRW Phonetics and Phonology of Speaking Styles, Barcelona, 19-1 a 19-5.
- Escudero, David/Aguilar, Lourdes/Vanrell, Maria del Mar/Prieto, Pilar (2012), *Analysis of inter-transcriber consistency in the Cat_ToBI prosodic labeling system*, *Speech Communication*, 54, 566-582.
- Eskénazi, Maxine (1993), *Trends in Speaking Styles Research*, in Proceedings of Eurospeech '93, Berlin, 501-505.
- Falé, Isabel (1995), *Fragmento da Prosódia do Português Europeu: as Estruturas Coordenadas*, Dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Falé, Isabel (2005), *Percepção e reconhecimento da informação entoacional em Português Europeu*, Dissertação de Doutoramento, Universidade de Lisboa.
- Falé, Isabel/Faria, Isabel (2006), *Categorical perception of intonational contrasts in European Portuguese*, in: Ralph Hoffman/Hans

- Mixdorff (Edd.), *Proceedings of Speech Prosody*, Dresden: TUDpress Verlag der Wissenschaften GmbH, 69-72,.
- Falé, Isabel/Faria, Isabel (2007), *Imperatives, orders and requests in European Portuguese intonation*, in: *Proceedings of the 16 International Congress of Phonetic Sciences*, 1041- 1044.
- Freitas, Maria João (1990), *Estratégias de organização temporal do discurso em português*, Dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Frota, Sónia (2000), *Prosody and Focus in European Portuguese, Phonological Phrasing and Intonation*, New York, Garland Publishing.
- Frota, Sónia (2002), *Nuclear falls and rises in European Portuguese: a phonological analysis of declarative and question intonation*, *Probus*, 14, 113-146.
- Frota, Sónia/D'Imperio, Mariapaola/Elordieta, Gorka/Prieto, Pilar/Vigário, Marina (2007), *The phonetics and phonology of intonational phrasing in Romance*, in: Pilar Prieto/Joan Mascaró/Maria-Josep Solé (edd.), *Prosodic and Segmental Issues in (Romance) Phonology*, Amsterdam/Philadelphia, John Benjamins, 131-153.
- Frota, Sónia (2014), *The intonational phonology of European Portuguese*, in Sun-Ah Jun (Ed.) *Prosodic typology II*, Oxford, Oxford University Press, 6-42.
- Frota, Sónia/Cruz, Marisa /Fernandes-Svartman, Flaviane/Collischonn, Gisela/Fonseca, Aline/Serra, Carolina/ Oliveira, Pedro/Vigário, Marina (2015), *Intonational variation in Portuguese: European and Brazilian varieties*, in: Sónia Frota/Pilar Prieto (edd.) *Intonation in Romance*. Oxford, Oxford University Press, 235-283.
- Grabe, Esther (1998), *Comparative Intonational Phonology: English and German*, Dissertação de Doutoramento, Universidade de Nijmegen. Publicada por MPI.
- Grabe, Esther/ Post, Brechtje/Nolan, Franci/Farrar, Kimberley (2000), *Pitch accent realisation in four varieties of British English*, *Journal of Phonetics* 28, 161-185.
- Grice, Martine/Benzmüller, Ralf/Savino, Michelina/Andreeva Bistra (1995), *The intonation of queries and checks across languages: data from map tasks dialogues*, in *Proceedings ICPHS 95*, Stockholm, 649-651.
- Grønnum, Nina/Viana, Maria do Céu (1999), *Aspects of European Portuguese intonation*, in: *Proceedings of ICPHS 99*, San Francisco, 1997-2000.
- Hirschberg, Julia (2000), *A corpus-based approach to the study of speaking style*, in: M. Horne (Ed.), *Prosody: theory and experiment*, *Festschrift in honor of Gösta Bruce*, Kluwer, 335-350.
- Julião, Mariana/Silva, Jorge/Aguiar, Ana/Moniz, Helena/Batista, Fernando (2015), *Speech features for discriminating stress using branch and bound wrapper search*, in: *SLATE'15*, Springer, series CCIS, Madrid, Spain.
- Jun, Sun-Ah (2005), *Prosodic Typology: The Phonology of Intonation and Phrasing*, Oxford University Press.
- Ladd, Robert (1996), *Intonational Phonology*, Cambridge, Cambridge University Press.

- Liu, Yi/Shriberg, Elizabeth/Stolcke, Andreas/Dustin, Hillard/Ostendorf, Mari/Harper, M. (2006), *Enriching Speech Recognition with Automatic Detection of Sentence Boundaries and Disfluencies*, IEEE Transactions on Audio, Speech, and Language Processing, 14, 1526-1540.
- Margolis, Anna (2011), *Automatic annotation of spoken language using out-of-domain resources and domain adaptation*, Dissertação de Doutorado, Universidade de Washington.
- Mata, Ana Isabel (1990), *Questões de entoação e interrogação no Português. Isso é uma pergunta?* Dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Mata, Ana Isabel (1999), *Para o estudo da entoação em fala espontânea e preparada no Português Europeu: metodologia, resultados e implicações didáticas*, Dissertação de Doutorado, Universidade de Lisboa.
- Mata, Ana Isabel/Santos, Ana Lúcia (2010), *On the intonation of confirmation seeking requests in child-directed speech*, in: Mark Hasegawa-Johnson (Ed.) Proceedings of Speech Prosody 2010, Chicago, U.S.A..
- Mata, Ana Isabel/Moniz, Helena/Batista, Fernando (2014), *Comparing phrase-final patterns across speech styles and age groups in European Portuguese*, *Nouveaux Cahiers de Linguistique Française* 31, 171-176.
- Mata, Ana Isabel/Moniz, Helena/Móia, Telmo/Gonçalves, Anabela/Silva, Fátima/Batista, Fernando/Duarte, Inês/Oliveira, Fátima/Falé, Isabel (2014a), *Prosodic, syntactic, semantic guidelines for topic structures across domains and corpora*, in: Nicoletta Calzolari/Khalil Choukri/Thierry Declerck/Hrafn Loftsson/Bente Maegaard/Joseph Mariani/Asuncion Moreno/Jan Odijk/Stelios Piperidis (Edd.) Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland 1188-1193.
- Mata/Ana Isabel/Moniz, Helena/Batista, Fernando/Hirschberg, Julia (2014b), *Teenage and adult speech in school context: building and processing a corpus of European Portuguese*, in: Nicoletta Calzolari/Khalil Choukri/Thierry Declerck/Hrafn Loftsson/Bente Maegaard/Joseph Mariani/Asuncion Moreno/Jan Odijk/Stelios Piperidis (Eds.) Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, 3914-3919.
- Mata, Ana Isabel/Moniz, Helena/Batista, Fernando, (2016), *Stylistic variation in the intonation of European Portuguese teenagers and adults*, in: Meghan Armstrong/Nicholas Henriksen/Maria del Mar Vanrell (edd.), *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields*, Amsterdam, John Benjamins, 45-68.
- Meinedo, Hugo/Neto, João Paulo (2003), *Audio segmentation, classification and clustering in a broadcast news task*, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03).

- Miller, Tim/ Stone, Dan (2009), *Public speaking apprehension (psa), motivation, and affect among accounting majors: A proof-of-concept intervention*, *Issues in Accounting Education*, 24, 265-298.
- Moniz, Helena (2006), *Contributo para a caracterização dos mecanismos de (dis)fluência em português europeu*, Dissertação de Mestrado, Faculdade de Letras, Universidade de Lisboa.
- Moniz, Helena/Trancoso, Isabel/Mata, Ana Isabel (2010), *Disfluencies and the perspective of prosodic fluency*, in: Anna Esposito/Nick Campbell/Carl Vogel/Amir Hussain/Anton Nijholt (edd.), *Development of multimodal interfaces: active listening and synchrony*, Berlin/Heidelberg, Springer-Verlag, 382-396.
- Moniz, Helena/Batista, Fernando/Trancoso, Isabel/Mata, Ana Isabel (2011), *Analysis of interrogatives in different domains*, in: Anna Esposito/Antonietta Esposito/Raffaele Martone/Vincent Müller/Gaetano Scarpetta (edd.), *Towards Autonomous, Adaptive, and Context-Aware Multimodal Interfaces: Theoretical and Practical Issues*, Berlin/Heidelberg, Springer-Verlag, 136-148.
- Moniz, Helena (2013), *Processing Disfluencies in European Portuguese*, Dissertação de Doutoramento, Universidade de Lisboa.
- Moniz, Helena/Mata, Ana Isabel/Hirschberg, Julia/Batista, Fernando/Rosenberg, Andrew/Trancoso, Isabel (2014a), *Extending AuToBI to prominence detection in European Portuguese*, in: Nick Campbell/David Gibbon/Daniel Hirst (edd.), *Proceedings of Speech Prosody 2014*, Dublin, Ireland, 280-284.
- Moniz, Helena/Batista, Fernando/Mata, Ana Isabel/Trancoso, Isabel (2014b), *Speaking style effects in the production of disfluencies*, *Speech Communication*, 65, 20-35.
- Moniz, Helena/Batista, Fernando/Mata, Ana Isabel/Trancoso, Isabel (no prelo), *Towards automatic language processing and intonational labeling in European Portuguese*, in: Meghan Armstrong/Nicholas Henriksen/Maria del Mar Vanrell (edd.), *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields*, Amsterdam, John Benjamins, 227-248.
- Moutinho, Lurdes/Coimbra, Rosa (2014), *Variation prosodique dans les interrogatives totales du Portugais Europeen continental*, in: Yolanda Congosto/ Maria Luísa Montero/Antonio Salvador (edd.), *Fonética Experimental, Educación Superior e Investigación*, vol. III, Madrid, Editorial Arco/Libros, 153-170.
- Moutinho, Lurdes/Coimbra, Rosa (2015), *As ondas da fala através do Atlântico. Um estudo acústico*, in: António Manuel Ferreira/Maria Fernanda Brasete (edd.), *Pelos Mares da Língua Portuguesa 2*, Aveiro, Universidade de Aveiro, 529-537.
- Neto, João Paulo/Meinedo, Hugo/Amaral, Rui/Trancoso, Isabel (2003), *A system for selective dissemination of multimedia information resulting from the ALERT project*, in: *Proceedings of the 2003 ISCA Workshop on Multilingual Spoken Document Retrieval (MSDR 2003)*, 25-30.
- Ostendorf, Mari/Favre, Benoît/Grishman, Ralph/Hakkani-Tür, Dilek/Harper, Mary/Hillard, Dustin/Hirschberg, Julia/Ji, Heng/Kahn, Joseph/Liu, Yang/Makey, S./Matusov, Evgeny/Ney,

- Herman/Rosenberg, Andrew/Shriberg, Elizabeth/Wang, Wen/Wooters, Chuck (2008), *Speech segmentation and spoken document processing*, IEEE Signal Processing Magazine, 25 (3), 59-69.
- Pierrehumbert, Janet (1980), *The Phonology and Phonetics of English Intonation*, Dissertação de Doutorado, Massachusetts Institute of Technology. Publicada em 1988 por IULC.
- Pitrelli, John/Beckman, Mari/Hirschberg, Julia (1994), *Evaluation of prosodic transcription labeling reliability in the ToBI framework*, *Proceedings of the 3rd International Conference on Spoken Language Processing, ICSLP 94*, Yokohama, vol. 1, 123-126.
- Price, Patti/Ostendorf, Mari/Shattuck-Hufnagel, Stephanie/Fong, Cynthia (1991), *The use of prosody in syntactic disambiguation*, JASA 90 (6), 2956-2970.
- Rodrigues, Celeste (no prelo), *Variação sociolinguística*, in: Ana Maria Martins/Ernestina Carrilho (edd.), *Manual de Linguística Portuguesa*, Berlin, De Gruyter.
- Santos, Vinícius/Fernandes-Svartman, Flaviane (2014), *O padrão entoacional neutro do português de Guiné-Bissau: uma comparação preliminar com o português brasileiro*. Estudos Linguísticos, São Paulo, 43 (1), 48-63.
- Silverman, Kim/Beckman, Mary/Pitrelli, John/Ostendorf, Mari/Wightman, Colin/Price, Patti/Pierrehumbert, Janet/Hirschberg, Julia (1992), *TOBI: a standard for labeling English prosody*, in: J. J. Ohala/T. M. Nearey/B. L. Derwing/M. Hodge/G. E. Wiebe (edd.), *Proceedings of ICSLP 92*, Banff, Canada, vol. 2, 867-870.
- Trancoso, Isabel/Viana, Maria do Céu/Duarte, Inês/Matos, Gabriela (1998), *Corpus de Diálogo CORAL*, in: *PROPOR'98, Actas do III Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada*, Porto Alegre, Brasil.
- Trancoso, Isabel/Martins, Rui/Moniz, Helena/Mata, Ana Isabel/Viana, Maria do Céu (2008), *The LECTRA Corpus - Classroom Lecture Transcriptions in European Portuguese*, in: Nicoletta Calzolari/Khalid Choukri/Bente Maegaard/Joseph Mariani/Jan Odijk/Stelios Piperidis/Daniel Tapias (edd.), *Proceedings LREC'08*, Marrakech, Morocco, ELRA, 1416-1420.
- Vaissière, Jacqueline (1983), *Language-independent prosodic features*, in: Anne Cutler/ D. Robert Ladd (edd.), *Prosody: Models and Measurements*, Berlin, Springer-Verlag, 53-66.
- Viana, Maria do Céu (1987), *Para a Síntese da Entoação do Português*, Dissertação da Carreira de Investigação, Universidade de Lisboa.
- Viana, Maria do Céu/Frota, Sónia/Falé, Isabel/Fernandes, Flaviane/Mascarenhas, Isabel/Mata, Ana Isabel/Moniz, Helena/Vigário, Marina (2007), *Towards a P_ToBI*, apresentação no *Workshop on the Transcription of Intonation in Ibero-Romance*, Universidade do Minho, Portugal. link
- Vigário, Marina (2003), *The prosodic word in European Portuguese*, Berlin, Mouton De Gruyter.
- Vigário, Marina/Frota, Sónia (2003), *The intonation of Standard and Northern European Portuguese: a comparative intonational*

- phonology approach*, Journal of Portuguese Linguistics 2-2 (special issue on Portuguese Phonology edited by W. L. Wetzels), 115-137.
- Wells, John C. (1997), *SAMPA computer readable phonetic alphabet*, in: Dafydd Gibbon/Roger Moore/Richard Winski (edd.), *Handbook of Standards and Resources for Spoken Language Systems*, Berlin/New York, Mouton de Gruyter, Part IV, section B.
- Wennerstrom, Ann (2001), *The music of everyday speech: prosody and discourse analysis*, New York, Oxford University Press.

ANA ISABEL MATA E HELENA MONIZ