Adaptação acústico-prosódica local em Português Europeu

Vera Cabarrão[1, 2], Fernando Batista[1, 3], Helena Moniz[1, 2, 4], Isabel Trancoso[1, 5]& Ana Isabel Mata[2]

[1]L2F, INESC-ID, Lisboa, [2]Universidade de Lisboa, CLUL, [3]ISCTE, Instituto Universitário de Lisboa

[4]Unbabel Lda, Lisboa, [5]IST, Universidade de Lisboa

Abstract:

This paper presents an acoustic-prosodic analysis of entrainment in map-task dialogues in European Portuguese. Our main goal is to analyze how turn-by-turn entrainment varies with distinct structural metadata events: types of sentence-like units (SUs) in consecutive turns (e.g. interrogatives followed by declaratives, or both declaratives), and with the presence of discourse markers, affirmative cue words, and disfluencies in the beginning of turns. Results show that entrainment at turn-exchanges occurs in terms of pitch, energy, duration, and voice quality. Considering SUs types, question-answer pairs are the ones with stronger similarity, as declarative-interrogative turns are the ones where less entrainment occurs. Moreover, in question-answer pairs, Yes/No and Tag questions present stronger evidences of entrainment than Wh-questions. Regarding turn-initial structures, there are evidences of (i) stronger entrainment when the second turn begins with an affirmative cue word, (ii) less strong with ambiguous structures (such as 'OK'), emphatic affirmative answers, and negative cue words; (iii) and scarce with disfluencies and discourse markers. Different degrees of local entrainment may be related to the informative structure of distinct structural metadata events.

***Keywords*:** local entrainment, acoustic-prosodic features, dialogues

Palavras-chave: adaptação local, parâmetros acústico-prosódicos, diálogos

1. Introdução

A adaptação entre falantes, do inglês *entrainment*, corresponde à capacidade que os seres humanos possuem de ajustar o seu comportamento e discurso ao do seu interlocutor, de forma a estabelecer uma interação de sucesso (Brennan *et al*., 1996; Benus, 2014). Estas estratégias de ajuste têm sido estudadas sob diversas perspetivas: a nível acústico-prosódico (Levitan e Hirschberg, 2011; Levitan, 2014; Gravano *et al*., 2014), lexical e sintático (Pardo, 2006; Nenkova *et al*., 2008; Lopes *et al*., 2013), psicológico e social (Benus, 2014), bem como a nível computacional (Levitan, 2014).

Estudos recentes sobre adaptação entre falantes em diferentes situações comunicativas têm mostrado a importância desta estratégia na resolução de tarefas específicas (e.g., diálogos em formato *map-task*, jogos colaborativos) e em sessões de terapia (e.g., terapia conjugal; sessões de aconselhamento sobre droga e linhas SOS anti-suicídio).

Estudos prévios sobre fala espontânea em Português Europeu (PE) mostraram que existe adaptação acústico-prosódica entre falantes a nível do diálogo (adaptação **global**), embora em diferentes graus: os falantes não se adaptam sempre aos mesmos interlocutores e seguindo as mesmas estratégias acústico-prosódicas (Cabarrão *et al*., 2016). O presente trabalho visa adicionar mais um nível à análise da adaptação acústico-prosódica em PE, especificamente observar como os falantes se adaptam a nível **local**, ou seja, enunciado a enunciado. O principal objetivo deste estudo é verificar se a adaptação entre pares de enunciados contíguos produzidos por falantes distintos varia de acordo com os **tipos de frase** (como, por exemplo, uma declarativa seguida de uma interrogativa ou quando ambas são declarativas ou interrogativas), bem como com a presença das seguintes estruturas no início do segundo enunciado de outro falante: **marcadores discursivos**

(e.g., *agora*; *bem*; *bom*; *portanto*; *então*), **disfluência**s, nomeadamente pausas preenchidas (e.g., *aa* e *aam*); **repetições enfáticas** (e.g., *sim, sim, sim*), **constituintes afirmativos** (e.g., *sim*; *exato*; *certo*) e **negativos** (e.g., *não*; *eu não tenho*) e **estruturas ambíguas** (palavras que tanto podem ser marcadores discursivos como constituintes afirmativos, como *pronto* e *ok*). Estas estruturas são muito frequentes em fala espontânea, em geral, e nos dados analisados, em particular (Batista, 2011; Batista *et al*., 2012b; Cabarrão, 2013; Moniz, 2013; Moniz *et al*., 2015; Cabarrão *et al*., 2016).

Este artigo está organizado da seguinte forma: a Secção 2 apresenta uma breve revisão bibliográfica de estudos sobre adaptação acústico-prosódica entre falantes; a Secção 3 descreve os dados, as pistas prosódicas e as métricas utilizadas; a Secção 4 apresenta os resultados obtidos relativamente à adaptação local entre falantes e às experiências efetuadas com diferentes tipos de frases e com estruturas específicas no início do enunciado contíguo de outro falante. Finalmente, a Secção 5 apresenta as conclusões obtidas e direções de trabalho futuro.

2. Enquadramento teórico

Este estudo visa não só analisar a adaptação local entre falantes, ou seja, enunciado a enunciado, mas também observar se esta depende do tipo de frases e das estruturas que ocorrem no início do segundo enunciado de outro falante, nomeadamente marcadores discursivos, disfluências, constituintes afirmativos e negativos.

A adaptação entre falantes, quer a nível do diálogo (global), quer entre enunciados contíguos (local), tem sido amplamente estudada em várias línguas (e.g., inglês, mandarim, eslovaco), mas não particularizando estruturas linguísticas frequentes em interações diádicas. Desta forma, até onde sabemos, não foi encontrada literatura sobre adaptação prosódica entre diferentes tipos de frase e poucos estudos se centram na adaptação entre falantes com estruturas específicas, como marcadores discursivos ou constituintes afirmativos/negativos.

A análise da estratégia de adaptação em diálogos não é recente (Grice, 1975; Giles *et al*., 1987, 1991; Brennan e Clark, 1996). Na Teoria da Acomodação, Giles *et al*. (1991) postulam que os falantes ajustam dinamicamente os seus comportamentos de comunicação, convergindo ou divergindo do seu interlocutor, para diminuir ou aumentar a distância entre si, respetivamente.

Em estudos recentes, a adaptação não é analisada de forma independente, mas tendo em conta as suas implicações face a certos objetivos, seja o sucesso do diálogo (Nenkova *et al*., 2008; Reitter e Moore, 2014), variáveis sociais (Chartrand, 1999; Benus, 2012), como, por exemplo, um falante ser considerado mais flexível e inteligente pelo seu interlocutor, ou em relações de poder (Danescu, 2012; Benus, 2014). De um ponto de vista linguístico, esta estratégia tem sido estudada sob diversas perspetivas: a nível acústico-prosódico (Levitan e Hirschberg, 2011; Levitan, 2014; Gravano *et al.*, 2014), fonético-fonológico (Pardo, 2006) e léxico-sintático (Ward, 2007; Nenkova *et al.*, 2008; Lopes, 2013). Os estudos apresentados nesta secção centram-se particularmente na perspetiva acústico-prosódica, tópico deste trabalho.

Brazil (1985) descreve como um falante adapta a frequência fundamental da sua voz à do interlocutor, expressando, assim, concordância com o que foi dito. Este fenómeno, designado pelo autor como concordância de *pitch* (*pitch concord*), é apresentado mais tarde em Whichmann (2000), Wichmann (2012) e Wennerstrom (2001), também como estratégia para resolver situações de conflito entre falantes. As relações de poder entre os participantes de um diálogo tornam-se evidentes ao identificar o falante que adapta a frequência fundamental ao seu interlocutor e aquele que mantém o seu registo habitual.

Mais recentemente, Levitan *et al*. (2011) e, posteriormente, Levitan (2014) mediram a adaptação entre falantes a nível global e local no *Columbia Games Corpus* (Gravano et al., 2009), um *corpus* de jogos colaborativos em inglês americano. Globalmente, os autores observaram que os falantes são mais semelhantes aos seus interlocutores do que aos não interlocutores, ou seja, falantes com quem não interagem, nos parâmetros média e máximo de energia e débito de fala. Localmente, os falantes assemelham-se mais nos

enunciados adjacentes do que nos não adjacentes nos parâmetros média e máximo de energia e *Harmonics-to-Noise-Rate* (HNR, medida da proporção entre um som harmónico e ruído na voz, em dB), mas não no débito de fala. Com base no trabalho de Edlund *et al*. (2009), em que os autores propuseram um modelo para medir a similaridade entre interlocutores suecos à medida que o diálogo progride, Levitan (2014) também analisou a sincronia (correlação entre os valores relativos de dois enunciados contíguos). Os resultados mostraram uma sincronia positiva em alguns parâmetros (sendo a média de intensidade o mais saliente), bem como sincronia negativa na maioria dos parâmetros, uma vez que os falantes não se ajustavam em relação aos seus interlocutores, mas no sentido oposto. Além disso, em Levitan (2014), a autora também encontrou evidências de adaptação nas palavras que sinalizam tomada de palavra (*turn-taking cues*, como *ok*), na medida em que um falante tende a utilizar as mesmas palavras do seu interlocutor. Adicionalmente, estes também registam mais palavras em comum entre si do que quando comparados com outros falantes. Heldner *et al*. (2010) já tinham mostrado que a frequência fundamental de um falante que produz um *backchannel* (quando um falantes sinaliza ao outro que está a acompanhar o discurso e que este pode continuar a falar) é mais similar ao do enunciado imediatamente anterior do que a qualquer outra palavra do diálogo.

Parcos são ainda os estudos que têm analisado a adaptação acústico-prosódica tendo em conta as funções pragmáticas de palavras específicas produzidas pelos falantes, nomeadamente pausas preenchidas, constituintes afirmativos e marcadores discursivos.

Benus *et al.* (2011) analisou padrões de acomodação temporal em estratégias de tomada de palavras entre falantes (o objetivo era estabelecer conhecimento partilhado pelos falantes - *common ground*). Os autores afirmam que a produção de pausas preenchidas em posição inicial (e.g., *um* ou *uhm*) e de respostas afirmativas só com um constituinte (e.g., *mmhm*, *ok* ou *yes*) está relacionada com funções pragmáticas específicas, como sinalizar relações de poder entre interlocutores, compreender conhecimentos partilhados e acomodar-se ao discurso do interlocutor. Os resultados mostraram que as relações de domínio entre falantes eram estabelecidas pelo controlo temporal, ou seja, pelo tempo limitado que um falante permitia ao seu interlocutor começar ou terminar o seu discurso. Pelo contrário, a acomodação estava relacionada com várias estratégias para ajustar o alinhamento temporal de pausas preenchidas e de respostas só com um constituinte afirmativo entre falantes.

Em Benus *et al.* (2012), os autores estudaram especificamente a adaptação de pausas preenchidas (e.g., *uh; hm; eh*) produzidas por juízes e advogados durante alegações do Supremo Tribunal de Justiça americano. O objetivo era analisar se a frequência e a qualidade das pausas preenchidas produzidas pelos pares advogado-juízes variava de acordo com o voto dos últimos. Os autores calcularam a diferença entre o número de vezes que um tipo de pausa preenchida era usado pelos falantes em cada interação. Os pares foram previamente divididos em dois grupos diferentes, aqueles que votaram a favor e aqueles que votaram contra. Globalmente, os resultados não mostraram influência na frequência e no tipo de pausas entre pares nos votos obtidos. Porém, a nível local, observou-se uma correlação significativa entre a qualidade das pausas preenchidas e a direção dos votos, na medida em que, quando um advogado produzia pausas preenchidas semelhantes à dos juízes, havia uma tendência para que o voto destes favorecesse o caso do advogado.

Benus (2014) analisou a adaptação entre falantes com o uso do marcador discursivo 'no' (que significa *sim* em eslovaco). Tal como acontece em PE (Cabarrão, 2013), em eslovaco, este marcador também contém funções pragmáticas distintas, como *backchannel*, concordância com o que foi dito ou mesmo início de um novo segmento discursivo. Os resultados mostraram menos adaptação do que o esperado na frequência de 'no' entre interlocutores: apenas os diálogos com um falante específico mostraram uma ocorrência elevada deste marcador. Em relação à sua função pragmática, a função *backchannel* ("reconheço que compreendo, por favor, continue") foi a mais frequente. O autor mostrou ainda que os falantes tendem a mostrar adaptação entre si quando ocorre este marcador, mas também no restante diálogo, nos parâmetros energia e qualidade de voz.

Até ao momento, o trabalho de Benus (2014) é o único a correlacionar adaptação acústico-prosódica com um marcador discursivo. No presente estudo, será realizada uma análise holística da adaptação acústico-

prosódica, tendo em conta diversas estruturas do PE, nomeadamente marcadores discursivos, disfluências (especificamente pausas preenchidas), constituintes afirmativas e também tipos de frase.

Em PE, apenas três estudos analisaram adaptação acústico-prosódica entre falantes (Cabarrão *et al*., 2013 2016a, 2016b), sendo que todos utilizaram o mesmo *corpus* aplicado no presente trabalho. Cabarrão et al. (2013; 2016a) mostraram evidências de correlações prosódicas (efeitos de concordância de níveis de frequência fundamental - f_0) entre perguntas sim/não e respetivas respostas afirmativas. Em Cabarrão et al. (2016b e 2018a), os autores mostraram evidências de adaptação nos diálogos (adaptação global), embora expressa em diferentes graus. Os falantes mostraram-se mais semelhantes aos seus interlocutores do que ao seu próprio discurso noutros diálogos na maioria dos parâmetros acústico-prosódicos, sendo a energia o único parâmetro inalterado. Já na comparação entre pares e não pares, a adaptação entre pares verificou-se maioritariamente em parâmetros de duração. Este estudo mostra que todos os parâmetros prosódicos (f_0, energia, duração, qualidade de voz) são monitorizados no processo de adaptação entre falantes, evidenciando um resultado para o PE que se diferencia dos obtidos para outras línguas (inglês e mandarim - Levitan et al., 2011; Levitan, 2014; Xia *et al*., 2014). Adicionalmente, os autores observaram que a adaptação entre falantes se relacionava mais com o interlocutor, a sua postura e personalidade, do que com o papel que este desempenhava no diálogo, de dador ou de seguidor.

Por outro lado, vários estudos já se centraram na análise de propriedades acústico-prosódicas dos diferentes tipos de frase em PE, bem como de algumas das estruturas em causa, nomeadamente, marcadores discursivos e disfluências.

No que diz respeito aos tipos de frase em PE, as SUs delimitadas com pontos finais (declarativas) estão associadas a contornos nucleares baixo-descendentes (por exemplo, Viana, 1987; Falé, 1995; Cruz-Ferreira, 1998; Frota, 2000), e Viana et al. (2007) associa as declarativas neutras ao contorno H+L* L%. Quanto às frases interrogativas, as parciais (Wh-) são caracterizadas com um contorno entoacional descendente, semelhante às declarativas neutras (Cruz-Ferreira, 1998; Viana *et al*., 2007); as perguntas sim-não são caracterizadas na fala espontânea em PE por Mata (1990) com contornos baixo-descendentes ou baixo-ascendentes; e por Falé (1995), com o contorno H* HL* H%. Em dados laboratoriais, Frota (2000) caracterizou-as com o contorno H+L* LH%. É de salientar também que este subtipo não possui pistas léxico-sintáticas associadas em português, mas somente prosódicas, ao contrário do que acontece para o inglês (interrogativas codificadas com um verbo auxiliar e inversão de sujeito). Quanto às *Tag*, Cruz-Ferreira (1998) descreve-as com contornos descendentes, Mata (1990) associa-as a uma melodia baixo-ascendente e em Mata e Moniz (2016), estas são associadas ao contorno L*+H H%.

Nos dados utilizados neste estudo, a recuperação automática de marcas de pontuação e respetiva análise acústico-prosódica foram efetuadas por Batista *et al*. (2007), Batista et al. (2012a), Batista et al. (2012b) e Moniz (2013). Os autores observaram que a identificação de marcas de pontuação e também de disfluências (*e.g.*, pausas preenchidas lexicalizadas, como "aam" e/ou "mm", apagamentos, substituições, entre outros) nas transcrições automáticas do reconhecedor de fala (Neto *et al*., 2008) permitiu uma melhoria significativa do *output* do sistema, o que resultou numa diminuição da taxa de erro de reconhecimento.

Quanto às disfluências, estas foram amplamente estudadas em PE por Moniz (2006; 2013) e Moniz *et al*. (2012; 2014; 2015). Estas estruturas são caracterizadas principalmente por: (i) duas palavras contíguas idênticas; tanto a energia como a frequência fundamental aumentam na palavra seguinte com um contorno *plateau* na palavra anterior; e (iii) um maior nível de confiança associado à palavra seguinte do que à anterior.

Finalmente, no que diz respeito aos marcadores discursivos, Cabarrão *et al*. (2016; 2018b) efetuaram uma tarefa de identificação destas estruturas no *corpus* usado neste estudo. Os autores apenas estudaram os marcadores que tendem a ocorrer na fala espontânea, são desprovidos de conteúdo proposicional e que, como tal, podem ser substituídos por uma disfluência, por exemplo. Na literatura (por exemplo, Lease e Johnson, 2006; Goldwater *et al*., 2010), os marcadores discursivos são frequentemente comparados às disfluências, uma vez que compartilham algumas funções pragmáticas, como manter a palavra enquanto se planeia o que dizer a seguir.

A caracterização prosódica das disfluências e dos marcadores discursivos enquanto classes independentes ainda não é consensual. A literatura aponta para várias estratégias: (i) considerar pausas preenchidas e reformulações como um subtipo da classe marcadores discursivos, ainda que defendendo que estes têm uma natureza distinta; ou (ii) considerar as disfluências como uma classe independente dos marcadores discursivos. Ao comparar os dois eventos nos mesmos *corpora*, Cabarrão *et al*. (2018b), após uma tarefa de classificação automática para distinguir marcadores discursivos, disfluências e SUs, concluíram que, de facto, os marcadores que têm uma função pragmática semelhante às disfluências partilham com estas características acústico-prosódicas, pelo que será legítimo considerá-las como parte de uma mesma classe.

3. Metodologia

3.1. *Corpus*

Este estudo utilizou o *corpus* CORAL[1] (ISLRN 499-311-025-331-2, Trancoso *et al.*, 1998), que contém 64 diálogos em formato *map-task* entre 32 falantes (equilibrados em termos de género). Estes não têm contacto visual e apenas podem comunicar oralmente. O grau de familiaridade entre interlocutores varia desde pares que não se conhecem a um par de gémeas idênticas.

Os diálogos ocorrem entre dois falantes que desempenham diferentes papéis, um é o dador de informação e o outro, o seguidor. Todos os participantes desempenham o papel de dador e de seguidor duas vezes com interlocutores diferentes. O dador dispõe de um mapa com uma rota e alguns marcos e tem como tarefa guiar o seguidor, para que este reconstrua o mesmo caminho no seu mapa incompleto. Os mapas foram desenhados de forma a desencadear estratégias de resolução colaborativas. O *corpus* tem 7 horas de fala ortograficamente transcritas, com um total de 61 mil palavras.

A amostra utilizada neste estudo corresponde a um conjunto de 48 diálogos entre 24 falantes (12 homens e 12 mulheres). Esta amostra está dividida em constituintes similares a frases (*sentence-like units* - SUs), com um total de 42 mil palavras. De acordo com Batista *et al.* (2012), as SUs podem corresponder a uma frase gramatical ou a uma unidade semântica menor do que uma frase.

3.2. Anotação do *corpus*

Como referido anteriormente, este trabalho visa observar se há adaptação acústico-prosódica entre falantes a nível local (enunciado a enunciado) e se essa adaptação ocorre com determinados tipos de frase e com estruturas específicas no início do segundo enunciado.

Quanto aos tipos de frase em enunciados contíguos, foram considerados os seguintes pares:

1. duas frases declarativas contíguas, com um total de 2770 ocorrências (DECL-DECL);
2. uma frase interrogativa seguida de uma declarativa, o típico par pergunta-resposta, com 838 ocorrências (INT-DECL);
3. uma frase declarativa seguida de uma interrogativa, com 676 casos (DECL-INT);
4. uma frase declarativa seguida de uma exclamativa, com 84 ocorrências (DECL-EXCL);
5. duas frases interrogativas contíguas, com 61 casos (INT-INT).

Os pares com frases exclamativas a anteceder ou a suceder interrogativas foram excluídos desta análise devido ao escasso número de dados (5 e 1 ocorrências, respetivamente).

Adicionalmente, também se procurou aferir se os diferentes tipos de interrogativas influenciam a adaptação entre pares pergunta-resposta. Como tal, as interrogativas nos pares INT-DECL foram classificadas com os seguintes subtipos:

[1] O corpus CORAL está disponível no ELRA *Catalogue of Language Resources*, com a referência ELRA - S0367.

1. Sim-não (*Estás a ver um túnel?*), com um total de 495 ocorrências;
2. Parciais (*Como é que é esse forte?*), 125 ocorrências;
3. *Tags* (*Está à esquerda dos cavalos selvagens, não é?*), 174 ocorrências.

As interrogativas alternativas foram excluídas desta análise dado o reduzido número de casos, 35 ocorrências (*e.g. Do lado esquerdo ou do lado direito?*). Adicionalmente, 9 casos foram também excluídos, uma vez que as questões não se integravam em nenhum destes subtipos de interrogativas (e.g., *Desculpa*?; *Então?*).

Relativamente às estruturas que ocorrem no início do enunciado contíguo, foram consideradas as seguintes: marcadores discursivos (DMs), constituintes afirmativos (ACW), estruturas ambíguas (AMB) - palavras que podem ser um DMs ou um ACW e disfluências (DF). Além disso, também foram consideradas outras estruturas muito frequentes nos dados utilizados, nomeadamente a repetição enfática (EMP) e os constituintes negativos (NEG).

A Tabela 1 lista todos os eventos analisados, respetivos exemplos e percentagens de ocorrências.

Tipos de frase			Estruturas em posição inicial		Exemplos
DECL-DECL		7%	DM	13%	*agora ; bem/bom ; portanto ; então*
INT-DECL	Sim-Não (60%)	2%	ACW	44%	*sim ; exacto/exactamente ; certo ;* grunts *(humhum* e *hum);* forma fixa do verbo "ser
	Tag (15%)		EMP	5%	*sim, sim, sim*
	Parciais (21%)		NCW	8%	*não ; eu não tenho*
DECL-INT		2%	AMB	18%	*pronto* 'ok'; ok
INT-INT		0.2%	DF	13%	pausas preenchidas *aa; aam*

Tabela 1: Tipos de frase e estruturas no início do segundo enunciado.

3.3. Parâmetros prosódicos e métricas

As experiências para medir o grau de adaptação entre falantes foram efetuadas com base no conjunto de parâmetros acústico-prosódicos eGeMAPS (*Extended Geneva Minimalistic Acoustic Parameters for Voice Research and Affective Computing*), apresentada por Eyben (2016). As eGeMAPS correspondem a pistas tipicamente adotadas em tarefas paralinguísticas e consistem num conjunto de 88 parâmetros que incluem: pistas de f_0, energia e parâmetros de voz (*jitter*, *shimmer* – medidas de qualidade da voz) com valores absolutos (designadas por Eyben *et al*., 2010, de descritores de baixo nível) e valores funcionais extraídos a partir dos valores absolutos (estatísticas, regressão polinomial, transformadas).

Para a análise de adaptação local, foram comparadas as características acústico-prosódicas entre o final de um enunciado, produzido por um falante, com o início do seguinte, produzido pelo seu interlocutor (ver exemplo em baixo). De seguida, foram aplicadas as métricas (Equações 1 e 2) propostas por Levitan e Hirschberg (2011) e Levitan (2014). Estas são baseadas em unidades inter-pausais (*Inter-Pausal Units*, IPU), ou seja, unidades de fala sem silêncio de um único falante separadas por, pelo menos, 50 ms (Gravano, 2009; Levitan, 2014). Na Equação 1, é calculada a semelhança entre o IPU final de um enunciado produzido por um falante e o IPU inicial do enunciado contíguo produzido pelo seu interlocutor. Na Equação 2, é calculada a semelhança entre o IPU final de um falante e o inicial de 10 enunciados não contíguos do seu interlocutor, escolhidos aleatoriamente no mesmo diálogo. Para se verificar se há adaptação local, as semelhanças entre enunciados contíguos têm de ser maiores do que entre enunciados não contíguos.

Exemplo: Dador: IPU SILÊNCIO IPU SILÊNCIO **IPU**
Seguidor: **IPU** SILÊNCIO IPU

$$(1) \quad PartnerDistance = |IPU_t - IPU_p|$$

$$(2) \quad OtherDistance = \frac{\sum |IPU_t - IPU_i|}{10}$$

Ao aplicar estas métricas, foi necessário ajustar a unidade de análise para dar conta dos fenómenos fonológicos do PE, como truncamentos de material pós-tónico, africadas ou aspiração. Outra razão para ajustar a unidade de análise foi a delimitação das estruturas-alvo (em posição inicial de enunciado). Assim, em vez de selecionar a IPU inicial e final para cada enunciado, foram selecionadas as palavras inicial e final produzidas dentro de um intervalo de 500 ms. Esta unidade mínima fixa de análise foi empiricamente testada e provou ser a medida mais adequada ao incluir uma ou mais palavras por unidade de análise. Tal permitiu extrair marcadores discursivos e constituintes afirmativos, que podem ser uma única palavra ou corresponder à combinação de duas ou mais palavras (por exemplo, "sim, sim, sim"; "sim, eu tenho"). Esse intervalo também pode ser usado para facilitar a classificação automática das estruturas-alvo e para produzir modelos de adaptação para sistemas automáticos de diálogo.

Os exemplos em baixo ilustram a classificação aplicada quanto aos tipos de frase e estruturas em início de enunciado, bem como ao alvo desta análise, a comparação entre o final de um enunciado de um falante com o início do enunciado contíguo do seu interlocutor.

Exemplos: Dador: Então, estás a ver uma loja de **chapéus**? (INT Sim/Não)
Seguidor: **Sim**, estou a ver a loja de chapéus. (DECL_AFF)

Dador: E chegas **ao fim**. (DECL)
Seguidor: **Ok**, já cá estou. (DECL_AMB)

Os valores obtidos foram depois comparados com um *paired t-test*, de forma a determinar: (i) se as semelhanças são maiores entre enunciados contíguos ou não (*Partner Distance vs. Other Distance*); (ii) considerando apenas os enunciados contíguos, se os falantes são mais semelhantes entre si quando o enunciado ocorre entre tipos de frase específicos ou quando o segundo enunciado começa com uma determinada estrutura. Este teste estatístico permite verificar a existência de diferenças estatisticamente significativas entre os valores obtidos de cada equação. Após determinar se há diferenças estatisticamente significativas entre os enunciados contíguos vs. não contíguos, a polaridade do valor *t* indica em que grupo as semelhanças entre falantes são maiores ou menores (um valor negativo indica que dentro dos enunciados contíguos há maiores semelhanças entre falantes, enquanto um valor positivo indica que essa semelhança é maior nos não contíguos). Assim, ao comparar os dois grupos, é possível observar em qual deles e em que parâmetros os falantes são mais semelhantes entre si.

4. Resultados

A presente secção dará conta da distribuição dos resultados relativos à adaptação local, bem como ao impacto, nessa adaptação, dos tipos de frase e de determinadas estruturas no início de um enunciado contíguo

de um interlocutor. Por questões de legibilidade, não é possível apresentar os resultados para todos os parâmetros acústico-prosódicos analisados, tendo sido feita uma seleção dos mesmos em função da sua significância.

4.1. Adaptação local

Em Cabarrão *et al*. (2016a; 2016b 2018a), os autores encontraram evidências de adaptação acústico-prosódica entre falantes por diálogo (adaptação global), no mesmo *corpus* em formato *map-task* do presente estudo. Agora, o principal objetivo é verificar se os mesmos falantes também mostram semelhanças entre si, mas entre enunciados contíguos produzidos por falantes diferentes (adaptação local).

Os resultados obtidos mostram diferenças estatisticamente significativas entre enunciados contíguos e não contíguos em 85 dos 88 traços acústico-prosódicos analisados (alguns desses parâmetros estão representados na Tabela 2). Estes resultados reforçam os obtidos no estudo sobre adaptação global em PE, uma vez que os falantes se assemelham aos seus interlocutores entre enunciados contíguos nos quatro parâmetros prosódicos: *f0*, energia, duração e qualidade da voz. Globalmente, os falantes mostram adaptação em apenas três pistas: declive de *f0*, duração de fala, com e sem silêncio interno, e rácio de fonação. Estes resultados diferem dos obtidos para o inglês americano e o mandarim (Levitan, 2014), em que os falantes mostram adaptação nos parâmetros energia (médio e máximo de energia) e HNR, mas não em parâmetros de f_0. Portanto, pode colocar-se aqui a hipótese de a adaptação com parâmetros relacionados com energia, mas não com pistas de f_0, poder ocorrer independentemente da língua, pelo menos em *corpora* semelhantes.

Parâmetros		t
f o	f0_amean	-36.6
	f0_pctlrange0_2	-49.4
	f0_meanRisSlope	-29.0
	f0_meanFallSlope	-24.5
	slopeV0_500_amean	-48.5
	slopeV500_1500	-51.6
	slopeUV0_500	-48.8
	slopeUV500_1500	-48.7
Energia	loudness_amean	-51.9
	loudness_pctlrange0_2	-49.7
	loudness_meanRisSlope	-51.0
	loudness_meanFallSlope	-48.2
	loudnessPeaksPerSec	-54.7
Qualidade de voz	jitter_amean	-48.8
	shimmer_amean	-55.0
	HNR_amean	-45.5
Vozeamento/ Desvozeamento	VoicedSegmentsPerSec	-49.7
	MeanVoicSegLengthSec	-61.0
	MeanUnvoiSegLength	-31.1

Tabela 2: T-tests (df=4449): *partner distance* vs. *other distance* (adaptação local). Todos os parâmetros apresentam significância estatística (p<0,001).

A mesma análise (enunciados contíguos vs. não contíguos) foi realizada para cada um dos 48 diálogos individualmente, tendo-se observado também evidências de adaptação na maioria dos parâmetros analisados. Ainda assim, um dos diálogos entre as duas irmãs gémeas (diálogo 30; s24-s21) difere dos restantes. Neste não foram observadas diferenças significativas na maioria dos parâmetros acústico-prosódicos. Este resultado já era expectável, dada a curta duração da interação, com apenas 4 enunciados contíguos entre os interlocutores. A maioria das SUs foi produzida pelo mesmo falante (s24), com instruções sobre o melhor caminho para resolver o mapa.

Tendo ainda em conta as semelhanças entre falantes encontradas por Cabarrão *et al*. (2016a; 2016b), também se procurou aferir neste estudo se os falantes que mostram um grau de adaptação maior a nível global também mostram adaptação local. Assim, foram analisados com mais pormenor os 10 falantes que, globalmente, apresentam semelhanças com o mesmo interlocutor independentemente do papel desempenhado, ainda que nem sempre nos mesmos parâmetros.

Os resultados mostram que os falantes são mais semelhantes entre si nos enunciados contíguos do que nos não contíguos. Esta semelhança é significativa ($p < 0{,}001$ e $p < 0{,}05$) na maioria dos parâmetros analisados, particularmente em f_0, energia, qualidade de voz e segmentos vozeados/desvozeados (ver Tabela 3). Novamente, o par s24, como dador, e s21, como seguidor, mostram diferentes resultados. Ainda que s24 apresente mais semelhanças com s21 do que com o outro falante com quem interage (s9), não há SUs suficientes para realizar uma análise enunciado a enunciado, visto a sintonia entre irmãs gémeas ser de tal ordem que a duração do diálogo é assaz reduzida.

G	s1	s17	s4	s8	s6	s18	s14	s22	s21	s24
F	s17	s1	s8	s4	s18	s6	s22	s14	s24	s21
Parâmetros	t (df 45)	t (df 42)	t (df 78)	t (df 70)	t (df 122)	t (df 99)	t (df 92)	t (df 89)	t (df 38)	t (df 3)
f0_amean	-5.99	-6.04	-4.6	-5.31	-10.53	-9.64	-6.35	-7.58	-2.11	**-1.94**
f0_pctlrange0_2	-6.12	-6.14	-8.16	-8.16	-9.03	-8.68	-9.65	-8.43	-5.83	**0.29**
f0_meanRisSlope	-3.72	-3.4	-3.78	-4.07	-6.24	-6.95	-5.4	-4.32	-3.72	**-2.13**
f0_meanFallSlope	-4.06	**-1.6**	-3.29	-4.54	-6.67	-4.78	-7.51	-6.93	-4.1	**-2.41**
slopeV0_500_amean	-6.05	-5.36	-7.21	-6.27	-9.82	-7.85	-6.87	-8.27	-3.857	**-0.79**
slopeV500_1500	-5.44	-6.16	-7.29	-6.99	-7.88	-7.71	-9.07	-5.97	-5.046	**-2.76**
slopeUV0_500	-6.59	-5.21	-7.4	-6.49	-8.41	-5.43	-6.99	-8.18	-4.957	**-0.91**
slopeUV500_1500	-6.64	-4.4	-6	-7.11	-8.41	-6.54	-7.62	-7.19	-5.706	**-1.21**
loudness_amean	-4.32	-5.71	-8.69	-6.67	-8.14	-7.27	-8.04	-10.12	-4.06	**0.12**
loudness_pctlrange0_2	-4.98	-6.8	-8.12	-7.76	-9.09	-7.29	-7.31	-9.79	-3.7	**0.44**
loudness_meanRisSlope	-5.74	-6.23	-8.09	-6.68	-8.06	-7.62	-7.68	-9.08	-5.94	**-0.51**
loudness_meanFallSlope	-5.86	-3.96	-9	-6.04	-8.25	-8.45	-6.54	-9.71	-4.27	**-1.28**
loudnessPeaksPerSec	-5.19	-5.61	-8.6	-7.55	-9.48	-8.58	-6.02	-8.18	-5.59	**-2.58**
jitter_amean	-4.34	-6.87	-7.85	-7.52	-10.74	-8.24	-7.21	-8.99	-5.82	**1.19**
shimmer_amean	-5.42	-6.85	-7.97	-7.97	-10.28	-8.7	-7.5	-8.57	-5.04	**0.7**
HNR_amean	-5.89	-5	-5.58	-6.22	-10.37	-9.01	-7.64	-8.97	-3.46	**-2.12**
VoicedSegmentsPerSec	-5.56	-5.36	-7.03	-6.05	-8.04	-8.27	-6.99	-6.8	-4.8	**0.22**
MeanVoicSegLengthSec	-6.72	-4.68	-8.21	-7.8	-11.1	-9.07	-9.47	-9.99	-5.2	-5.51
MeanUnvoiSegLength	-4.08	-5.92	-4.62	-5.89	-9.68	-8.8	-8.25	-8.06	-5.74	-3.24

Tabela 3: T-tests (df=4449): *partner distance* vs. *other distance* (adaptação local). Todos os parâmetros apresentam significância estatística ($p < 0{,}05$ e $p < 0{,}001$), exceto os marcados a negrito.

4.2. Adaptação local com diferentes estruturas

A adaptação local foi observada em termos de f_0, energia, duração e qualidade de voz. Além disso, já havia evidências de que os falantes tendem a adaptar-se a interlocutores específicos, independentemente do papel desempenhado e do género (Cabarrão *et al*., 2016, 2018a). Ainda assim, não é claro se a adaptação local também depende dos tipos de frases, das estruturas linguísticas no início do enunciado contíguo do interlocutor e das suas funções informativas. Como tal, nesta secção, será analisado se a adaptação local varia com diferentes tipos de SUs em enunciados contíguos (por exemplo, interrogativos seguidos de declarativos ou ambos declarativos) e com a presença de marcadores discursivos, constituintes afirmativos e disfluências no enunciado contíguo do interlocutor. O principal objetivo é aferir se existem graus de adaptação de acordo com estas estruturas específicas, similarmente aos graus encontrados ao nível do papel desempenhado e do interlocutor na adaptação global.

4.2.1. Análise da adaptação local com diferentes tipos de frase

O *corpus* de diálogo varia entre enunciados interrogativos e declarativos, correspondendo a perguntas sobre a localização de alguns pontos de referência no mapa ou informações ou direções a seguir, respetivamente. Nos dados, há também ocorrências de enunciados exclamativos, geralmente interjeições (para mais informações sobre os sinais de pontuação no corpus CORAL, ver Batista *et al*., 2007, 2010; Moniz *et al*., 2010).

Os dados usados para analisar a adaptação local contêm 3621 frases declarativas, 738 interrogativas e 89 exclamativas. Para investigar se a adaptação difere de acordo com os tipos de SUs em enunciados consecutivos, foi efetuada a mesma análise aplicada para testar a adaptação local (*Partner Distance vs. Other Distance*, conferir secção 3). O objetivo é verificar se os enunciados contíguos são mais semelhantes do que os não contíguos, considerando cada tipo de SUs.

Tendo em conta os enunciados declarativos (DECL) e interrogativos (INT), os resultados mostram que os falantes são mais semelhantes entre si nos contíguos do que nos não contíguos nos quatro parâmetros prosódicos: frequência fundamental, energia, duração e qualidade de voz. No entanto, um teste de Kruskall-Wallis, comparando apenas os enunciados contíguos para estes tipos de frase, também revela que existem diferenças estatisticamente significativas entre eles ($p<0,001$ e $p<0,05$) na maioria das pistas acústico-prosódicas. Tal mostra que, embora os enunciados contíguos sejam sempre mais parecidos do que os não contíguos, também diferem de acordo com o tipo de frase.

Para verificar em que SUs os falantes são mais semelhantes, realizou-se um *paired t-test* a comparar os diferentes padrões entre dois enunciados consecutivos (ver secção 3), tendo-se contabilizado o número de parâmetros (de um total de 88) em que cada par se assemelha mais entre si. Devido ao número reduzido de enunciados exclamativos, em comparação com os declarativos e os interrogativos, estes SUs não foram considerados na análise.

A Tabela 4 mostra os resultados obtidos. Cada célula representa o rácio de parâmetros acústico-prosódicos com significância estatística ($p<0,001$ e $p<0,05$) em que os falantes são mais semelhantes, para cada combinação de tipos SUs. **Os resultados mostram que os falantes são mais semelhantes nos pares pergunta-resposta (INT-DECL) do que entre DECL-DECL (25/5) ou entre DECL-INT (40/15)**. Em ambas as comparações, os pares pergunta-resposta são mais semelhantes nos quatro parâmetros prosódicos: frequência fundamental, energia, duração e qualidade de voz. Nos pares DECL-DECL também são encontradas evidências de adaptação quando comparados com os pares DECL-INT: no primeiro par, os falantes mostram semelhanças em 35 parâmetros e, no segundo, em apenas 10. Na comparação entre pares pergunta-resposta com os pares INT-INT, os resultados são menos expressivos, na medida em que menos pistas apresentam semelhanças significativas entre os falantes (4/1). Quanto à comparação entre DECL-DECL vs. INT-INT e DECL-INT vs. INT-INT, os resultados são muito equilibrados, não havendo uma clara tendência de que os falantes de um par são mais semelhantes que no outro.

	INT-INT	DECL-INT	DECL-DECL
INT-DECL	4/1	**40/15**	**25/5**
DECL-DECL	0/1	**35/10**	
DECL-INT	5/6		

	Parciais-DECL	TAG-DECL
Sim/Não-DECL	6/5	14/10
TAG-DECL	13/7	

Tabela 4: Rácio de parâmetros acústico-prosódicos, de um total de 88, com significância estatística (p<0,001 e p<0,05), em que os falantes são mais semelhantes por tipos de SUs.

Em suma, os pares pergunta-resposta são os que apresentam maiores semelhanças entre os falantes, sendo os pares declarativo-interrogativo aqueles nos quais ocorre menos adaptação. Nos dados, observa-se que os enunciados declarativos geralmente correspondem a uma resposta a uma pergunta anterior ou a informações sobre a posição no mapa, seguida por uma pergunta sobre o próximo passo a realizar para a conclusão da tarefa. Este *corpus* em formato *map-task* é caracterizado pela sua natureza colaborativa, em que os falantes interagem com um objetivo comum, o de chegar ao destino o mais rapidamente possível. Assim sendo, era expectável que os pares pergunta-resposta fossem aqueles com maior adaptação.

Tendo em conta apenas os pares pergunta-resposta, observa-se a existência de graus de adaptação entre os diferentes tipos de interrogativas, embora as diferenças não sejam tão evidentes quanto o esperado. A comparação entre perguntas Sim-Não e Parciais, em que ambas são seguidas por uma resposta declarativa, mostra que estes pares apresentam diferenças estatisticamente significativas em 11 parâmetros: em 6 deles, os falantes são mais semelhantes quando há uma pergunta Sim-Não, e em 5 quando há uma Parcial. Quanto aos pares Sim-Não-DECL vs. TAG-DECL, os resultados também mostram que os falantes são mais semelhantes entre as perguntas Sim-Não e a seguinte resposta (14/10). As semelhanças ocorrem em frequência fundamental, energia e parâmetros espectrais. Finalmente, ao comparar TAG-DECL com WH-DECL, as evidências de adaptação são mais acentuadas no primeiro par (13/7). Estes resultados podem ser explicados pelo facto de que estes SUs são codificados prosodicamente de formas distintas.

4.2.2. Análise da adaptação local com diferentes estruturas em posição inicial

Esta análise foi realizada somente entre as frases interrogativas, nomeadamente, perguntas Sim-Não e *Tags* (como ambos mostraram resultados semelhantes, foram agrupados numa única classe, para se obter um maior número de ocorrências), seguido por uma resposta declarativa (Sim-Não/*Tag*-DECL), e pelos pares DECL-DECL. Esta seleção deveu-se ao facto de as estruturas alvo de análise não ocorrerem em número suficiente (menos de 20 casos) no início do segundo enunciado nos restantes pares (INT-INT, DECL-INT e WH-DECL). Como esperado, nos pares INT-DECL, os marcadores discursivos e as estruturas ambíguas também não ocorrem com muita frequência, 21 e 24 casos, respetivamente, pelo que também foram excluídos desta análise.

As Tabelas 5 e 6 mostram o rácio de parâmetros acústico-prosódicos com significância estatística (p<0,001 and p<0,05), em que os falantes são mais semelhantes. Nos pares pergunta-resposta, os resultados mostram que os falantes se adaptam mais quando a resposta começa com um constituinte afirmativo em vez de uma repetição enfática (13/4), nos parâmetros de frequência fundamental, *jitter* e HNR, ou de uma disfluência (15/6), principalmente em pistas de qualidade de voz e segmentos vozeados/desvozeados. Quanto aos constituintes afirmativos e negativos, os resultados são muito semelhantes, uma vez que os falantes

mostram evidências de adaptação num número similar de parâmetros (13/12). Relativamente aos pares DECL-DECL, os falantes também tendem a ser mais semelhantes aos seus interlocutores quando o enunciado contíguo começa com um constituinte afirmativo do que com uma repetição enfática (mais semelhanças em 20 parâmetros, principalmente em frequência fundamental e energia); marcadores discursivos (45/13), estruturas ambíguas (32/26), tanto em frequência fundamental, energia, parâmetros espectrais e características de qualidade de voz, e disfluências (31/12), em energia, qualidade de voz e segmentos vozeados/desvozeados. Contrariamente aos pares pergunta-resposta, os constituintes afirmativos e negativos não são balanceados em termos da quantidade de pistas em que os interlocutores mostram adaptação (19/11). Ao comparar marcadores discursivos com todas as outras estruturas analisadas, os resultados mostram que esta classe é aquela em que os falantes menos se adaptam. Quanto às disfluências, há evidências de maior semelhança entre interlocutores apenas quando comparadas com os marcadores (30/12).

	NCW	DFs	EMP
ACW	13/12	15/6	13/4
EMP	15/15	10/12	
DFs	12/13		

Tabela 5: Rácio de parâmetros acústico-prosódicos (com significância estatística de $p<0{,}001$ e $p<0{,}05$) em que os falantes são mais semelhantes, nos pares Sim-Não/*Tag*-DECL

	NCW	DFs	DM	AMB	EMP
ACW	19/11	**31/12**	**45/13**	32/26	**20/9**
EMP	16/16	**26/8**	**29/10**	12/22	
AMB	16/10	**22/7**	**49/10**		
DMs	**8/30**	**12/30**			
DFs	11/23				

Tabela 6: rácio de parâmetros acústico-prosódicos (com significância estatística de $p<0{,}001$ e $p<0{,}05$) em que os falantes são mais semelhantes, nos pares DECL-DECL

Nos pares em que o segundo enunciado começa com repetições enfáticas, estruturas ambíguas e constituintes negativos, os falantes mostram mais adaptação do que naqueles que começam com disfluências, na maioria dos parâmetros acústico-prosódicos. Portanto, há maiores evidências de adaptação quando o enunciado contíguo começa com um constituinte afirmativo, quer o enunciado anterior seja uma declarativa ou uma interrogativa; essa adaptação é menos evidente com estruturas ambíguas, repetições enfáticas e constituintes negativos; sendo pouco frequente com disfluências e marcadores discursivos. Estes diferentes graus de adaptação podem estar relacionados com a estrutura informacional destes eventos. Nos dados em causa, os constituintes afirmativos têm diversas funções pragmáticas, como expressar concordância ou apenas dar reforço positivo para o interlocutor continuar o seu discurso. Independentemente da sua função, estas estruturas contribuem para a fluidez do diálogo e sinalizam a natureza colaborativa do corpus. Por outro lado, quer os marcadores discursivos selecionados quer as disfluências são geralmente definidas como estruturas sem conteúdo proposicional, que compartilham propriedades acústico-prosódicas de acordo com seu contexto pragmático: os marcadores que têm uma função semelhante às disfluências, como planear as estruturas subsequentes, podem partilhar com elas alguns propriedades, ou seja, os contornos *plateau* que contrastam com o aumento de frequência fundamental no constituinte prosódico seguinte.

Em suma, nos dados em causa, a adaptação é influenciada pelos tipos de SUs e pelas estruturas que ocorrem no início do segundo enunciado: os falantes assemelham-se mais aos seus interlocutores nos pares pergunta-resposta (mais semelhanças num maior número de parâmetros do que em qualquer outro par de SUs analisado), bem como quando o segundo enunciado começa com um constituinte afirmativo (e não com marcadores discursivos ou disfluências).

5. Conclusão

Este estudo apresenta uma primeira análise de adaptação local (enunciado a enunciado) em PE. O principal objetivo era o de perceber se os tipos de frases em enunciados consecutivos (por exemplo, interrogativos seguidos por declarativos, ou ambos declarativos) e a presença de marcadores discursivos, constituintes afirmativos e disfluências no início do enunciado contíguo influenciam a adaptação acústico-prosódica entre os falantes.

Os resultados gerais obtidos quanto à adaptação local, i.e., sem considerar os tipos frases ou estruturas específicas, revelam que as produções dos falantes são mais semelhantes entre enunciados contíguos do que entre não contíguos nos quatro parâmetros prosódicos: frequência fundamental, energia, duração e qualidade de voz. Estes resultados não estão de acordo com uma análise similar realizada por Levitan (2014) para o inglês americano: os falantes adaptam-se nos enunciados contíguos nos parâmetros média e máximo de energia e HNR, mas não na frequência fundamental. Esses resultados permitem colocar a hipótese de que pistas como energia podem ser independentes da língua, pelo menos em corpora semelhantes, mas não os parâmetros de frequência fundamental. As experiências realizadas até ao momento mostram assim que o comportamento acústico-prosódico da adaptação local em PE se estende das pistas de energia a todos os outros parâmetros prosódicos.

Considerando a adaptação entre diferentes tipos de SUs, os pares pergunta-resposta são aqueles com maior semelhança na maioria dos parâmetros analisados, frequência fundamental, energia, duração e qualidade de voz, sendo os pares declarativo-interrogativo aqueles nos quais ocorre menos adaptação. Estes resultados já eram expectáveis, dada a natureza colaborativa da tarefa do *corpus*. Quanto aos subtipos de interrogativas nos pares pergunta-resposta, existem evidências mais fortes de adaptação com perguntas Sim-Não e *Tags* do que com perguntas parciais. Os dois primeiros compartilham um contorno alto/ascendente por oposição ao contorno baixo/descendente também associado às declarativas neutras em PE. Além disso, as perguntas Sim-Não não possuem pistas léxico-sintáticas associadas em português, apenas prosódicas, o que pode constituir mais uma evidência para a adaptação local encontrada. Importa ainda referir que os pares pergunta-resposta são a força motriz da natureza dialógica do *corpus*, constituído por tarefas muito colaborativas que têm de ser resolvidas em conjunto pelos dois interlocutores. A fluidez de um diálogo baseia-se em várias estratégias e os dados deste estudo mostram que as estruturas com maior adaptação local são as que promovem colaboração e reforço positivo.

Em linha com o que foi dito para as SUs, há maiores evidências de adaptação local com constituintes afirmativos no início do segundo enunciado em enunciados contíguos. Estes resultados evidenciam, uma vez mais, o esforço colaborativo entre os interlocutores para resolver a tarefa. Por outro lado, as disfluências e os marcadores discursivos são as estruturas que apresentam menor grau de adaptação. Uma possível explicação é o facto de que, quando os falantes proferem estas estruturas, estão a planear o que dizer a seguir. Os padrões de planeamento das estruturas subsequentes em PE são *plateaus* que se distinguem dos padrões prosódicos de outras estruturas linguísticas.

Num trabalho futuro, pretende-se realizar uma análise mais refinada das diferentes funções pragmáticas dos marcadores discursivos e dos constituintes afirmativos, de forma a verificar como estes se correlacionam com a adaptação entre falantes. Pretende-se, também, estender este estudo para outros domínios, como chamadas telefónicas em serviços de *call-center*.

Referências

Batista, Fernando, Diamantino Caseiro, Nuno Mamede & Isabel Trancoso (2007) Recovering punctuation marks for automatic speech recognition. In *Proceedings of Interspeech* 2007, pp. 2153 – 2156, Antuérpia, Bélgica.

Batista, Fernando (2011) *Recovering Capitalization and Punctuation Marks on Speech Transcriptions*. Tese de Doutoramento, Instituto Superior Técnico.

Batista, Fernando, Helena Moniz, Isabel Trancoso, & Nuno Mamede (2012a) Bilingual experiments on automatic recovery of capitalization and punctuation of automatic speech transcripts. *Transactions on Audio Speech and Language Processing*, (20), pp. 474–485.

Batista, Fernando, Helena Moniz, Isabel Trancoso, Nuno Mamede & Ana Isabel Mata (2012b) Extending automatic transcripts in a unified data representation towards a prosodic-based metadata annotation and evaluation. *Journal of Speech Sciences*, (3), pp.115–138.

Benus, Štefan, Agustín Gravano & Julia Hirschberg (2011) Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics* 43.12 (2011), pp. 3001-3027.

Benus, Stefan, Rivka Levitan & Julia Hirschberg (2012) Entrainment in spontaneous speech: the case of filled pauses in Supreme Court hearings. In *3rd IEEE Conference on Cognitive Infocommunications*.

Beňuš, Stefan (2014a) Social aspects of entrainment in spoken interaction. In *Cognitive Computation*, 6(4), pp. 802-813.

Benus, Stefan (2014b) Conversational entrainment in the use of discourse markers. In *Recent Advances of Neural Network Models and Applications*, pp. 345–352. Springer.

Brazil, David (1985) Phonology: Intonation in discourse. *Handbook of discourse analysis*, 2, pp. 57–75.

Brennan, Susan E. e Herbert H. Clark (1996) Conceptual pacts and lexical choice in conversation. In *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22 (6), pp. 1482-1493.

Cabarrão, Vera (2013) *Respostas afirmativas em diálogos espontâneos em português europeu: interface prosódia-sintaxe-discurso*. Tese de mestrado, Faculdade de Letras da Universidade de Lisboa

Cabarrão, Vera, Ana Isabel Mata e Isabel Trancoso (2016a) Affirmative constituents in european portuguese dialogues: prosodic and pragmatic properties. In *Proceedings of Speech Prosody*, pp. 634–638.

Cabarrão, Vera, Isabel Trancoso, Ana Isabel Mata, Helena Moniz & Fernando Batista (2016b) Global analysis of entrainment in dialogues. In IberSpeech 2016, Springer, vol. 10077, series Lecture Notes in Computer Science, pag. 215-223, doi: 10.1007/978-3-319-49169-1_21, *Advances in Speech and Language Technologies for Iberian Languages: Third International Conference*, IberSPEECH 2016, Lisboa, 23-25 de novembro, 2016.

Cabarrão, Vera, Helena Moniz, Fernando Batista, Isabel Trancoso & Ana Isabel Mata (2018a). Adaptação acústico-prosódica entre falantes. In *Revista da Associação Portuguesa de Linguística* (4), pp. 18-33.

Cabarrão, Vera, Helena Moniz, Fernando Batista, Jaime Ferreira, Isabel Trancoso & Ana Isabel Mata (2018b). Cross-domain analysis of discourse markers in European Portuguese. In *Dialogue & Discourse 9*, No 1, pp. 79-106.

Chartrand, Tanya L. & John A. Bargh. (1999) The chameleon effect: the perception–behavior link and social interaction. In *Journal of personality and social psychology* 76 (6), pp. 893-910.

Cruz-Ferreira, Madalena (1998) Intonation in European Portuguese. In Hirst, D. and Di Cristo, A., editors, *Intonation systems*, pag. 167–178. Cambridge: Cambridge University Press.

Danescu-Niculescu-Mizil, Cristian, Lillian Lee, Bo Pang & Jon Kleinberg (2012) Echoes of power: Language effects and power differences in social interaction. In *Proceedings of the 21st international conference on World Wide Web*, pp. 699-708.

Edlund, Jens, Julia Hirschberg & Mattias Heldner (2009) Pause and gap length in face-to-face interaction. In *Tenth Annual Conference of the International Speech Communication Association*.

Eyben, Florian, Klaus R. Scherer, Björn W. Schuller, Johan Sundberg, Elisabeth André, Carlos Busso, Laurence Y. Devillers *et al.* (2016) The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. In *IEEE Transactions on Affective Computing 7(2)*, pp. 190-202.

Falé, I. (1995) *Fragmentos da prosódia do português europeu: as estruturas coordenadas.* Master's thesis, University of Lisbon.

Frota, Sónia (2000) *Prosody and Focus in European Portuguese. Phonological Phrasing and Intonation.* Garland Publishing, New York.

Giles, Howard, Anthony Mulac, James J. Bradac, & Patricia Johnson (1987) Speech accommodation theory: The first decade and beyond. In *Annals of the International Communication Association* 10(1), pp. 13-48.

Giles, Howard, Nikolas Coupland & Justine E. Coupland (1991) Accommodation theory: Communication, context, and consequence. In *Contexts of accommodation: Developments in applied sociolinguistics* (1), pp.1-68.

Goldwater, S., Jurafsky, D., and Manning, C. D. (2010) Which words are hard to recognize? prosodic, lexical, and disfluency factors that increase speech recognition error rates. *Speech Communication*, 52(3), pp. 181–200.

Gravano, Agustin (2009) *Turn-taking and affirmative cue words in task-oriented dialogue*. Dissertação de doutoramento, Universidade da Columbia.

Gravano, Agustín, Štefan Beňuš, Rivka Levitan & Julia Hirschberg (2014) Three ToBI-based measures of prosodic entrainment and their correlations with speaker engagement. In *Spoken Language Technology Workshop* (SLT), IEEE, pp. 578-583

Grice, Paul (1975) Logic and conversation. In Maite Ezcurdia, & Robert J. Stainton (eds.) *The semantics-pragmatics boundary in philosophy*. Broadview Press, pp. 41-58.

Heldner, Mattias, Jens Edlund, and Julia Bell Hirschberg (2010) "Pitch similarity in the vicinity of backchannels. In *Eleventh Annual Conference of the International Speech Communication Association*.

Lease, M. and Johnson, M. (2006) Early deletion of fillers in processing conversational speech. In Proceedings of the *Human Language Technology Conference* of the NAACL, Companion Volume: Short Papers, pages 73–76. Association for Computational Linguistics.

Levitan, Rivka (2014) *Acoustic-prosodic entrainment in human-human and human-computer dialogue*. Dissertação de Doutoramento, Universidade da Columbia.

Levitan, Rivka & Julia Hirschberg (2011) Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Proceedings of Interspeech* 2011, pp. 3081-3084.

Lopes, Jose, Maxine Eskenazi & Isabel Trancoso (2013) Automated two-way entrainment to improve spoken dialog system performance. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8372-8376.

Mata, Ana Isabel (1990) *Questões de entoação e interrogação no Português. Isso é uma pergunta?* Master's thesis, University of Lisbon.

Mata, Ana Isabel & Helena Moniz. 2016. Prosódia, variação e processamento automático. In Ana Maria Martins & Ernestina Carrilho (eds.), *Manual de Linguística Portuguesa*. MRL Series, Mouton de Gruyter: 116-155. DOI: 10.1515/9783110368840-007.

Moniz, Helena (2006) *Contributo para a caracterização dos mecanismos de (dis)fluência no Português Europeu*. Master's thesis, University of Lisbon.

Moniz, Helena, Batista, F., Trancoso, I., and Mata, A. I. (2012) Prosodic context-based analysis of disfluencies. In *Interspeech 2012*, Portland, Oregon.

Moniz, Helena (2013) *Processing disfluencies in european portuguese.* PhD thesis, University of Lisbon.

Moniz, Helena, Fernando Batista, Ana Isabel Mata & Isabel Trancoso (2014) Speaking style effects in the production of disfluencies. In *Speech Communication* (65), pp. 20-35.

Moniz, Helena, Jaime Ferreira, Fernando Batista & Isabel Trancoso (2015) Disfluency detection across domains. In *The 7th Workshop on Disfluency in Spontaneous Speech (DiSS 2015)*. International Phonetic Association.

Nenkova, Ani, Agustin Gravano & Julia Hirschberg (2008) High frequency word entrainment in spoken dialogue. In *Proceedings of the 46th annual meeting of the association for computational linguistics on human language technologies*, pp. 169-172. Association for Computational Linguistics.

Neto, João, Hugo Meinedo, Márcio Viveiros, Renato Cassaca, Ciro Martins, and Diamantino Caseiro (2008) Broadcast news subtitling system in Portuguese. In *ICASSP 2008*, pag. 1561–1564.

Pardo, Jennifer S. (2006) On phonetic convergence during conversational interaction. In *The Journal of the Acoustical Society of America* 119(4), pp. 2382-2393.

Reitter, David & Johanna D. Moore (2007) Predicting success in dialogue. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 808-815.

Trancoso, Isabel, Maria do Céu Viana, Inês Duarte & Gabriela Matos (1998) Corpus de Diálogo CORAL. I*n PROPOR'98*, Porto Alegre, Brasil.

Viana, M. C. (1987) *Para a Síntese da Entoação do Português.* PhD thesis, University of Lisbon.

Viana, Maria do Céu, Sónia Frota, Isabel Falé, Flaviane Fernandes, Isabel Mascarenhas, Ana Isabel Mata, Helena Moniz & Marina Vigário (2007). Towards a P_ToBI. In *Workshop of the Transcription of Intonation in the Ibero-Romance Languages*, PaPI 2007, Minho, Portugal.

Wennerstrom, Anne. (2001) *The music of everyday speech: Prosody and discourse analysis.* Oxford University Press.

Whichmann, Anne (2000) *Intonation in text and discourse: Beginnings, middles and ends.* Longman.

Wichmann, Anne (2012) Prosody in context: the effect of sequential relationships between speaker turns. *Prosody and meaning*, pp. 239–270.

Xia, Zhihua, Rivka Levitan & Julia Hirschberg (2014) Prosodic Entrainment in Mandarin and English: A Cross-Linguistic Comparison. In *Proceedings of Speech Prosody,* pp. 65-69.

Ward, Arthur & Diane Litman (2007) Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora. In *Workshop on Speech and Language Technology in Education.*