

REVISÃO DE LITERATURA

# MODELOS PSICOLINGUÍSTICOS DE RECONHECIMENTO DE SONS DE FALA: O LUGAR DO ACENTO LEXICAL NO ACESSO LEXICAL DE L1 E DE L2

Amanda Post da SILVEIRA  

Departamento de Letras – Universidade Federal de São Carlos (UFSCar)

## RESUMO

Este trabalho traz uma contribuição crítica sobre os modelos de reconhecimento de sons de fala, mais especificamente sobre o papel (quando existe em tais modelos) do acento de palavra tanto no sistema cognitivo tanto de monolíngues quanto de bilíngues no acesso lexical em fala contínua. Argumentaremos que em muitos dos modelos clássicos de reconhecimento de palavra não é previsto o papel do acento, o que modelos mais recentes têm tentado dar conta. Porém, evidenciamos que não há um modelo de representação, processamento e reconhecimento de fala bilíngue que explique o papel do acento de palavra. Mostraremos possíveis caminhos de como essa implementação pode ser feita, sem, contudo, exaurir as possibilidades. Esse trabalho pretende também trazer uma contribuição para a psicolinguística do bilinguismo, uma vez que traz uma revisão crítica de modelos mais relevantes de linha conexionista - notadamente, o modelo TRACE, o Shortlist e Shortlist B -, tentando promover um debate entre partes que trabalhem nessa linha. Pretendemos apresentar mais discussões sobre a relação acento lexical e acesso lexical em futuros trabalhos que estão sendo publicados em sequência.

 OPEN ACCESS

## EDITORES

– Miguel Oliveira, Jr. (UFAL)  
– René Almeida (UFS)

## AVALIADORAS

– Michele Alves (UFAC)  
– Elena Ortiz-Preuss (UFG)

## DATAS

– Recebido: 12/08/2021  
– Aceito: 13/09/2021  
– Publicado: 30/09/2021

## COMO CITAR

SILVEIRA, Amanda Post da (2021). Modelos psicolinguísticos de reconhecimento de sons de fala: o lugar do acento lexical no acesso lexical de L1 e de L2. *Cadernos de Linguística*, v. 2, n. 4, e502.

#### ABSTRACT

This work makes a critical contribution on speech sound recognition models, more specifically on the role (when it exists in such models) of word stress in both monolingual and bilingual cognitive systems in lexical access in continuous speech. We will argue that in many of the classic word recognition models the accent role is not foreseen, something that more recent models have tried to account for. However, we show that there is no model of representation, processing and recognition of bilingual speech that explains the role of word stress. We will show possible ways of how this implementation can be done, without, however, exhausting the possibilities. This work also intends to contribute to the psycholinguistics of bilingualism, since it brings a critical review of demost relevant connectionist line models - notably the TRACE model, the Shortlist and the Shortlist B - trying to promote a debate between parties working in this line. We intend to present more discussions about the relation between lexical accent and lexical access in future works that are being published in sequence.

#### PALAVRAS-CHAVE

Modelos de Reconhecimento de Fala; Acento Lexical; L1; L2.

#### KEYWORDS

Models of Speech Recognition; Word Stress; L1; L2.

## INTRODUÇÃO

Apesar de o acento primário (doravante, acento) ser uma propriedade inerente de cada palavra, seu papel no reconhecimento de palavras ainda não é bem compreendido, mesmo para monolíngues. Não sabemos em detalhes como o acento lexical é codificado na representação de uma palavra, ou como e quando ela é recuperada durante o processamento de texto. Mesmo quando os idiomas não marcam o acento com diacríticos em suas representações ortográficas, os falantes monolíngues são capazes de recuperar informações de acentuação da palavra para sua língua nativa durante a leitura (PERRY, ZIEGLER & ZORZI, 2010), e os bilíngues podem identificar e produzir sílabas tônicas (acentuadas), em sua segunda língua, mesmo quando há muitos recursos acústicos de incompatibilidade de acentuação de palavra entre os idiomas (e.g. CUTLER, 1986; VAN HEUVEN, 1996).

Para compreender o papel da acentuação de palavras na recuperação de palavras de L1 e de L2, utilizamos estudos anteriores que indicam que as representações de acentos de palavras são construídas em associação com o léxico da primeira língua (L1) e da segunda língua (L2), e que são influenciados pelas distribuições de frequência de padrão lexical e de padrões acentuais em L1 e em L2. Também mostraremos que o acento da palavra pode estar representado tanto pela via sublexical quanto pela via lexical. No nível sublexical, a acentuação da palavra pode ser associada a representações segmentais, fonotáticas e silábicas, enquanto que no nível lexical, as representações fonológicas, ortográficas e, até certo ponto, semânticas e o contexto pragmático entram em jogo.

Neste estudo mostraremos como o acento lexical pode desempenhar um papel no reconhecimento da palavra falada. Como sabemos, o padrão específico de acento de uma palavra pode distinguir esta palavra de outra palavra. Por exemplo, as palavras *trustEE* e *obJECT* com acento na segunda sílaba têm um significado diferente de *TRUSTy* e *OBject* com acento na primeira sílaba. Além disso, como descreveremos mais adiante com mais detalhes, o acento de palavra está associado a diferenças na duração, amplitude e especificação dos fonemas. Essas diferenças presentes no sinal acústico que caracterizam o acento devem afetar a velocidade e a precisão do reconhecimento de palavras.

Nas seções subsequentes, consideraremos os processos de reconhecimento de palavras faladas de L1 e de L2, revisaremos vários modelos psicolinguísticos de reconhecimento de palavras, ressaltando que estes muitas vezes não especificam as relações condicionais entre sílabas, estrutura silábica, posição da sílaba e acento.

Este pretende ser o segundo estudo de uma sequência de estudos que estão sendo publicados para a compreensão da representação, processamento e produção do acento lexical de monolíngues e bilíngues (ver também Post da Silveira, 2021).

## 1. ACENTO LEXICAL NO RECONHECIMENTO DE PALAVRAS FALADAS EM L1

Quando os ouvintes processam uma palavra falada, independentemente de serem monolíngues ou bilíngues, eles devem de alguma forma lidar com a variabilidade no sinal acústico de fala e segmentar o sinal em unidades significativas menores. Durante a compreensão das palavras faladas, a segmentação das unidades sonoras (por exemplo, fonemas) e a determinação dos limites das palavras acontecem *on-line* (que significa simultaneamente neste contexto) com o desdobramento do sinal de fala (NORRIS et al., 1997). Os mecanismos de invariância perceptual corrigem as diferenças acústicas com base na coarticulação contextual na realização dos segmentos, permitindo uma ativação precisa de *cohorts* (listas) de palavras candidatas (MARSLEN-WILSON & WELSH, 1978; MARSLEN-WILSON & TYLER, 1980).

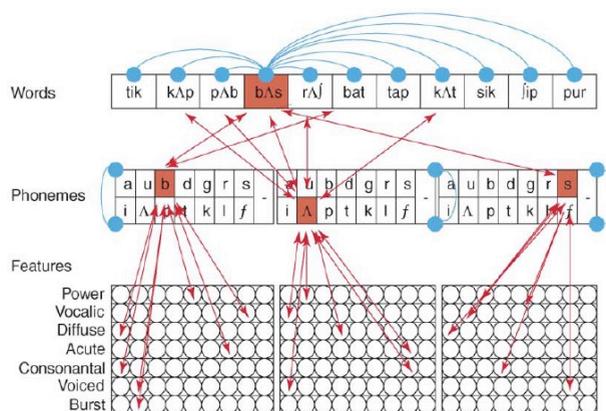
O reconhecimento de palavras faladas é desafiado não apenas pela necessidade de uma decodificação automática e precisa do sinal acústico, mas também pela ausência de limites lexicais no continuum da fala (NORRIS et al., 1997). Para facilitar o reconhecimento dos limites das palavras, os ouvintes fazem uso de padrões de silabificação (fonotática) e padrões rítmicos de sua língua nativa (L1). No caso de línguas com acento verbal, uma combinação de regularidades e consistências fonotáticas e silábicas auxilia no reconhecimento *on-line* de padrões sonoros que são mais prováveis de serem, por exemplo, as primeiras sílabas ou sílabas acentuadas de uma palavra. Em inglês, o padrão de acentuação mais frequentemente observado é o acento na primeira sílaba; portanto, as sílabas tônicas fornecem pistas para o início das palavras para ouvintes nativos da língua (CUTLER & NORRIS, 1988).

No entanto, a maioria dos modelos de reconhecimento de palavras faladas não considera os padrões de silabificação (fonotática) e padrões rítmicos de L1, do processo de reconhecimento de palavras. Eles especificam principalmente um conjunto de candidatos a palavras inicialmente ativados e sua redução a uma palavra-alvo com base no sinal de fala que se desdobra. Por exemplo, o modelo de *cohort* (MARSLEN-WILSON, 1978, 1990) propõe que o processo de reconhecimento da palavra falada começa no início acústico das palavras nos níveis sublexical e lexical. Assim que o início de uma palavra chega ao sinal de voz, os mecanismos de reconhecimento de palavras combinam o *input* acústico com as representações lexicais armazenadas, configurando um conjunto de possíveis candidatos a palavras. Isso é chamado de *cohort* inicial de palavra e consiste em todas as palavras que correspondem ao início do *input* acústico. O processamento continua *bottom-up* e da esquerda para a direita, seguindo o *input* falado: conforme o sinal de fala se desenvolve ao longo do tempo, os segmentos de fala subsequentes continuam a ativar o subconjunto do *cohort* que corresponde ao início do *input*, enquanto os membros do *cohort* não combinam,

não são mais pertinentes (e eliminados do *cohort* ou desativados). Frequentemente, as palavras podem ser reconhecidas antes de terminarem, porque podem ser eliminadas da ambiguidade de todos os outros candidatos com base no sinal de fala anterior. O ponto no sinal de voz onde há apenas um candidato único deixado *on-line* com esse sinal é chamado de ponto de exclusividade da palavra (MARSLEN-WILSON, 1980).

Na versão inicial do modelo de *cohort*, informações linguísticas de nível superior, como contexto semântico e de discurso, podiam exercer um efeito *top-down* no reconhecimento de palavras e acelerá-lo (cf. ZWITSERLOOD, 1989). Isso não é mais possível na versão posterior do modelo. Um problema relacionado é que o reconhecimento de palavras se torna problemático quando o início de uma palavra não é (totalmente) reconhecido: neste caso, um *cohort* inicial de palavras não pode ser configurado adequadamente e o contexto *top-down* não pode ajudar. A descrição do modelo não faz menção ao papel do acento na palavra na ativação dos membros de *cohort*; o reconhecimento de palavras, em princípio, ocorre com base apenas na identificação segmental.

Outro modelo influente para o reconhecimento de palavras faladas é o TRACE. Este é um modelo de ativação interativa conexionista (ELMAN & MCCLELLAND, 1988; MCCLELLAND & ELMAN, 1986; MCCLELLAND & RUMELHART, 1981; RUMELHART & MCCLELLAND, 1982), consistindo em níveis de representação característicos, fonêmicos e de palavras (ver Figura 1).



**Figura 1.** Arquitetura do modelo TRACE proposta por McClelland, Mirman, & Holt (2006). Ele contém conexões excitatórias bidirecionais (setas vermelhas) e conexões inibitórias (círculos azuis). Elementos consistentes em níveis adjacentes são mutuamente alimentados por excitação, enquanto dentro de camadas as unidades competem inibindo umas às outras (adaptado de McClelland et al., 2006).

As representações consistem em nós (unidades) em uma rede de representações que são ativadas na medida em que correspondem ao *input* acústico. Quando uma representação corresponde ao sinal acústico, ela é ativada e espalha a ativação para os nós conectados. Por exemplo, o fonema / p / ativará palavras das quais faz parte, como

/ pat / e / pack / em inglês. O fluxo de ativação é bidirecional: ele se espalha tanto dos níveis inferiores para os superiores (*bottom-up*) e vice-versa (*top-down*). As representações ativadas no mesmo nível se desativam, um mecanismo que é chamado de "inibição lateral". A ativação no modelo se espalha entre todos os nós e níveis simultaneamente. Para explicar o fato de que os fonemas podem retornar ao longo do tempo (como na palavra 'papai'), existem várias representações dependentes do tempo para cada fonema que se tornam ativas se o fonema (ou um similar) for detectado em uma janela de tempo específica. A ativação no nível da palavra é sensível à frequência lexical (ou seja, uma palavra de frequência mais alta tem um nível de ativação de repouso mais alto para começar) e ao tamanho do conjunto de competidores (ou seja, palavras em *cohorts* - conjuntos de competidores maiores que exercem e sofrem mais inibição lateral). TRACE difere do modelo de *cohort*<sup>1</sup>, porque pode se recuperar do problema de inícios de palavras não identificados, pois fonemas posteriores ainda podem ativar a palavra candidata adequada por *feedback*, assim efeitos *top-down* do nível lexical para o nível fonêmico podem resultar em uma espécie de restauração fonêmica parcial. Essa 'restauração' não poderia ser feita no nível do fonema, porque os fonemas reais podem ter sido alterados por informações *top-down* do nível lexical (ver Figura 1).

Potencialmente, por meio dos mecanismos de 'compensação para o contexto auditivo', 'adaptação seletiva' e 'sintonia de percepção da fala', como na implementação do modelo (MCCLELLAND et al., 2006), a restauração fonêmica poderia ocorrer com base na característica representação do sinal de voz de *input*. Essa representação foi inicialmente considerada como excluída das influências de nível superior em versões anteriores do modelo (ou seja, o conjunto de recursos é determinado com base no sinal de *input* e apenas envia informações *bottom-up* para o nível fonêmico; MCCLELLAND & ELMAN, 1986; ELMAN & MCCLELLAND, 1988). Os efeitos do contexto auditivo nos sons da fala ocorrem quando os segmentos vizinhos de uma palavra compartilham características acústicas ou articulatórias. Por exemplo, quando precedido por /j/, os fonemas /t/ e /k/ tenderão a ser ouvidos como /k/, porque /k/ é velar e /j/ é palatal. No entanto, precedidos de /s/, os fonemas /t/ e /k/ tendem a ser ouvidos como /t/, pois ambos compartilham a característica alveolar (exemplo de McClelland et al., 2006). A adaptação seletiva da fala pode ser explicada pela frequência de um som no léxico. Se um som for muito frequente, ele é facilmente desambiguado de outros sons menos frequentes que são semelhantes. Por

1 No início de uma palavra, tanto a Cohort como a TRACE prevêem que muitos itens léxicos são ativados em paralelo com os candidatos que abandonam esta ativação definido quando ocorre um descasamento no fluxo da fala, contudo diferem quanto ao processo de seleção de palavras. O modelo Cohort não incorporam conexões inibitórias *bottom-up* entre unidades fonêmicas e unidades lexicais que representam palavras que não contêm esses fonemas. Em vez disso, TRACE usa conexões inibitórias no nível léxico para descartar potenciais palavras candidatas. Por esta razão, o modelo só é capaz de reduzir a ativação de um item léxico após verificação de incongruências onde existem outras unidades mais ativas que podem fornecer inibição no nível léxico (Davis, 2000).

exemplo, se /s/ é altamente frequente e /ʃ/ não, /ʃ/ é facilmente eliminado de /s/. Se o componente interativo do TRACE estiver de acordo com a compreensão humana real, a adaptação seletiva deve afetar a identificação do fonema no nível de processamento sublexical, bem como os processos lexicais.

O modelo apresenta ainda mecanismos para ajustar o sistema aos dialetos e às diferenças individuais no *input* de fala. A previsão feita é que a variabilidade no input fonêmico pode ser resolvida tanto sublexicamente - por adaptação seletiva - quanto lexicamente - por consulta fonológica. No entanto, os processos subsequentes de identificação de palavras podem ser afetados como consequência: Erros de pronúncia em níveis fonêmicos e característicos podem ativar candidatos errados. Ao mesmo tempo, uma certa tolerância da variabilidade do sinal, conforme incluída no modelo, pode ajudar os ouvintes humanos a resolver o problema de invariância na percepção da fala.

Implicitamente, o modelo TRACE sustenta que a percepção humana mapeia o sinal acústico em características fonéticas que primeiro ativam unidades fonológicas (fonemas) e, em seguida, itens lexicais. TRACE I (MCCLELLAND & ELMAN, 1986) focou no processo inicial de mapeamento de propriedades acústicas da fala real (de vários falantes) em unidades fonêmicas abstratas. TRACE II foi criado para levar em conta as influências lexicais na percepção fonêmica (por exemplo, MCCLELLAND & RUMELHART, 1989; HOLT & LOTTO, 2010), em outras palavras, para testar a influência *top-down* do feedback lexical na percepção fonêmica. No TRACE II, o *input* de fala real foi substituída por fala simulada com a intenção de explorar os mecanismos responsáveis pelo papel de processamento *top-down* da informação lexical no reconhecimento de fonemas. Aqui, o nível lexical está envolvido desde o início do processo perceptivo, mas seus efeitos no nível fonêmico aumentam contínua e progressivamente, à medida que as próprias unidades lexicais recebem cada vez mais ativação com o desdobramento da palavra no sinal de fala ao longo do tempo. O modelo afirma que a força do feedback lexical é proporcional ao nível de ativação das unidades lexicais que contêm o fonema alvo. Assim, a hipótese é que as palavras podem fornecer *feedback* excitatório para os fonemas já antes que o reconhecimento de palavras ocorra. Essa proposição é pertinente ao momento em que os fonemas são codificados para o reconhecimento de palavras e faz a previsão de que *cohorts* de candidatos são ativados de forma *bottom-up* e eles próprios realimentam o processo de percepção fonêmica antes mesmo de ocorrer o reconhecimento de palavras. O reconhecimento de palavras ocorreria no "ponto de singularidade", o momento do sinal acústico em que a palavra-alvo se torna a única palavra que ainda é compatível com todos os fonemas apresentados até agora.

Como consequência do uso de fala simulada que consiste em fonemas com durações e características fixas no TRACE II, o problema de variância do reconhecimento de palavras humanas não é mais abordado de forma adequada. Ouvintes humanos ainda podem

recuperar uma palavra-alvo no caso de variação na realização fonética de fonemas que se desvia de qualquer categoria conhecida na língua. Eles são capazes de selecionar o candidato lexical que melhor corresponda ao *input*, mesmo se o sinal não for perfeito. Um modelo de reconhecimento auditivo de palavras como o TRACE não pode lidar adequadamente com esse aspecto da percepção humana.

Se os mecanismos de reconhecimento da fala humana exigiram uma precisão de "categoria restrita" no nível sublexical de decodificação fone-para-fonema para funcionar, a variabilidade na realização fonêmica tornaria a inteligibilidade impossível. Em vez disso, o 'melhor' candidato é provavelmente determinado comparando as características da palavra armazenada com as propriedades sublexicais do sinal de *input*. Assim, o mesmo mecanismo *top-down* que permitiria a "restauração" fonêmica no reconhecimento de palavras de L1, também pode ser responsável pela capacidade do monolíngue de reconhecer palavras estrangeiras com acento e da capacidade do ouvinte de L2 de reconhecer palavras em línguas estrangeiras envolvendo categorias fonêmicas que não fazem parte de seu sistema de L1.

O modelo TRACE foi aplicado a dados de reconhecimento de palavras de L1 com algum sucesso. Frauenfelder, Segui e Dijkstra (1990) realizaram uma série de experimentos de monitoramento de fonemas com fonemas-alvo em posições pontuais pré e pós-unicidade de palavras e não palavras que foram apresentadas auditivamente a ouvintes monolíngues de holandês e francês. Os fonemas-alvo que ocorrem após o ponto de exclusividade foram reconhecidos mais rapidamente do que os fonemas-alvo antes do ponto de exclusividade. Isso indica que processos de reconhecimento pós-lexical ocorrem e podem acelerar o reconhecimento fonêmico. No entanto, nenhuma evidência conclusiva foi obtida em favor da previsão do modelo TRACE de que a identificação fonêmica pode ser facilitada antes que o ponto de singularidade seja alcançado com base no feedback *top-down* do *cohort* lexical. Para esclarecer isso com um exemplo, quando o fonema / p / foi monitorado, a palavra holandesa "opera" conduziu a facilitação estatisticamente não significativa em relação ao "opelo", não-palavra pareada (em ambos os itens, o / p / alvo está situado logo antes do ponto de exclusividade). Os resultados mostraram que a informação lexical acelera a detecção de fonemas para fonemas após o ponto de singularidade das palavras portadoras do alvo; mas o efeito lexical não pode substituir informações *bottom-up* conflitantes, conforme previsto por TRACE.

Em um estudo de simulação adicional, Frauenfelder e Peeters (1998) observaram efeitos de gague facilitadores de palavras de *cohort* em fonemas-alvo via feedback *top-down* apenas para algumas palavras-alvo, enquanto outras não foram afetadas ou inibidas pelo mecanismo *top-down*. As adaptações do modelo mostraram que o modelo pode realmente reconhecer palavras mais rápido quando usa apenas um mecanismo ascendente de avanço. Não foram coletadas evidências experimentais suficientes no

momento para caracterizar completamente o mecanismo de *feedback* lexical no reconhecimento de fonemas. Propomos que a aplicação de técnicas sensíveis ao tempo, como o monitoramento on-line dos movimentos oculares, pode fornecer dados mais esclarecedores sobre o papel do feedback lexical no reconhecimento de fonemas antes do ponto de singularidade.

Assim, o modelo TRACE motivou *insights* a respeito da relação entre as representações sublexicais e lexicais durante o reconhecimento de palavras, mas os efeitos de acentuação da palavra não são levados em consideração. As diferenças de amplitude são, até certo ponto, levadas em consideração (em termos do traço de potência que caracteriza os fonemas), mas o acento silábico está ausente (ver Elman e McClelland (1988)). Portanto, consideraremos os desenvolvimentos mais recentes deste quadro teórico em termos da Lista Restrita A (*Shortlist A*) e da Lista Restrita B (*Shortlist B*).

No modelo TRACE, uma rede muito grande de aspectos, fonemas e representações lexicais é levada em consideração durante o reconhecimento de cada palavra de *input*. No entanto, a maioria dessas representações não são relevantes para a palavra em questão, o que torna todo o procedimento de reconhecimento bastante ineficiente (pelo menos em termos de simulações). Esse problema pode ser contornado formulando primeiro uma pequena lista de candidatos a palavras para cada *input* de palavra e, somente então, prosseguindo com o processo real de reconhecimento de palavras. Isso é o que é feito no modelo *Shortlist A* (NORRIS, 1994; NORRIS ET AL., 1997). Em primeiro lugar, as informações ascendentes ativam uma lista de candidatos a palavras com base em sua correspondência com o *input* acústico. Semelhante ao TRACE, um candidato a palavra é ativado com base em sua sobreposição com o sinal de *input* em diferentes posições na palavra. Apenas as palavras da lista restrita participam em seguida no processo de ativação interativo que leva ao reconhecimento. A próxima etapa envolve a segmentação da fala com base em inferência probabilística. Como a maioria das palavras consiste em pelo menos uma sílaba, o sinal de voz pode ser segmentado em unidades de tamanho de palavra, facilitando o reconhecimento de palavras com base no *input* de voz. Este procedimento, portanto, usa a restrição de palavras possíveis (Possible Word Constraint (PWC); NORRIS et al., 1997). Por exemplo, é mais difícil detectar “apple” (maçã) na sequência da não palavra “fapple” do que na sequência “vufapple”, porque, no último caso, uma palavra possível (em vez de uma única consoante) é deixada fora do *input* após a segmentação de “apple” (NORRIS et al., 1997).

Uma extensão da *Shortlist A* fornece uma melhoria para o PWC. A *Shortlist B* (NORRIS & MCQUEEN, 2008) mostra que muitos aspectos do reconhecimento de palavras faladas podem ser capturados por um modelo que deriva possíveis segmentos e combinações de segmentos em difonemas do sinal acústico com base em probabilidades de transição entre sílabas. Com relação à segmentação de palavras, a *Shortlist B* usa probabilidades de sílaba e informações métricas para prever os limites das palavras. Este algoritmo para

segmentação automática de fala é baseado em experimentos de localização de palavras em que falantes nativos de inglês acharam mais difícil localizar palavras como *sea* (mar) incorporadas em sequências como *seash* do que em sequências como *seashub* (NORRIS et al. 1997; NORRIS & MCQUEEN, 2008).

A dificuldade no primeiro caso surge da restrição de palavras possíveis (PWC) mencionada acima, ou seja, a preferência de dividir o sinal em unidades que são palavras possíveis (NORRIS, et al. 1997). O PWC foi complementado no modelo de segmentação *Shortlist B* por outro fator implicado na segmentação de sequências fonêmicas em palavras, a fonotática específica da linguagem. No exemplo de *seash*, o fonema /ʃ/ não pode ser uma sílaba nem uma palavra por si só e, portanto, os ouvintes têm dificuldade em aceitar *sea* independentemente de *-sh*. As duas partes não seriam fonotática ou metricamente aceitáveis na língua. Por outro lado, no exemplo de *seashub*, *shub* é uma sequência fonotática possível (CVC) e uma palavra potencial em inglês. Em suma, a fonotática específica do idioma e a lexicalidade das sílabas (a probabilidade de uma sílaba percebida ser uma palavra real) co-determinam o comportamento de segmentação da *Shortlist B* (para um teste deste procedimento de segmentação, consulte Vroomen & de Gelder, 1995).

Extensões mais recentes da *Shortlist B* (Norris & McQueen, 2008; Reinisch et al., 2010; Warren & Marslen-Wilson, 1987) afirmam que todas as informações acústicas recebidas são imediatamente usadas para o reconhecimento de palavras, incluindo a acentuação da palavra (SULPIZIO & MCQUEEN, 2012). Assim, as propriedades prosódicas e segmentares do sinal de voz são consideradas decodificadas em paralelo. Para dar conta do processamento pré-lexical da informação suprasegmental em paralelo com a informação segmentar (analisada por PWC), Sulpizio e McQueen (2012) propuseram um Analisador de Prosódia que extrai informações suprasegmentais do sinal de fala, para restringir simultaneamente candidatos ideais segmental e prosodicamente. Os autores ressaltam que o processamento segmental e suprasegmental é específico do idioma, e destacam que a acentuação da palavra em inglês é marcada segmentalmente pelo contraste de vogais completa versus reduzida que marca sílabas tônicas e átonas (não acentuadas), respectivamente.

Além disso, em um nível mais abstrato de representação lexical, o conhecimento sobre as frequências do padrão de acentuação das palavras no léxico é usado para modular o reconhecimento de palavras (SULPIZIO & MCQUEEN, 2012). No entanto, o acento lexical não é capaz de restringir o reconhecimento de palavras por si só. Sulpizio e McQueen (2012) argumentam que o léxico não é totalmente especificado no que diz respeito às características acústicas (ao contrário do que afirmam as teorias exemplares, por exemplo, Tenenhaus & Griffin, 2001; Pierrehumbert, 2002). Em vez disso, Norris e McQueen (2008) propõem que o reconhecimento da palavra falada é ótimo no sentido de que os ouvintes

lidam da melhor forma possível com as restrições impostas tanto pelo sinal de fala quanto por seu conhecimento fonológico e lexical. Como o sinal de fala não é perfeitamente claro, mas foneticamente ambíguo, e porque as palavras não são unidades discretas (têm um desdobramento quase contínuo), Norris e McQueen (2008) argumentam que um mecanismo de inferência bayesiana é ideal para o reconhecimento de palavras, porque combina a evidência perceptiva do sinal de fala, mesmo quando ambígua, com probabilidades lexicais aprendidas. Este mecanismo cognitivo faz uso de pistas acústicas segmentais e prosódicas aprendidas no reconhecimento de palavras (SULPIZIO & MCQUEEN, 2012) a fim de aprender categorias fonológicas abstratas para fonemas e acentos de palavras. Em contraste com o TRACE, as variantes e extensões da lista restrita representam modelos estritamente ascendentes de reconhecimento de palavra falada.

## 2. REPRESENTAÇÕES DE SONS DE FALA BILÍNGUE E ACESSO LEXICAL DE L2

Como vimos na seção anterior, a segmentação da fala é essencial no processo de reconhecimento da palavra falada de ouvintes nativos. As estratégias de segmentação demonstraram ser específicas da linguagem em relação aos padrões rítmicos predominantes na língua nativa (JUSCZYK, CUTLER, & REDANZ, 1993; CHRISTOPHE, MEHLER, & SEBASTIÁN-GALLÉS, 2001). Por exemplo, em idiomas não redutivos (ou com ritmo de sílaba), como espanhol, italiano e português brasileiro (PB), tanto a fonotática quanto os padrões de acentuação da palavra fornecem pistas confiáveis que podem auxiliar no reconhecimento de palavras (para uma análise do PB, consulte Post da Silveira et al., 2018).

Quando bilíngues processam a fala em uma segunda língua (L2), surgem novos problemas que não são uma preocupação no processamento monolíngue da fala. Em particular, os bilíngues possuem dois sistemas fonológicos interativos, um dos quais é dominante em bilíngues com dominância em uma das línguas (SEBASTIÁN-GALLÉS & KROLL, 2003). Embora a aquisição de novos sons contrastivos de L2 possa ocorrer com relativa facilidade, os sons que compartilham semelhanças entre L1 e L2 apresentam dificuldades de aprendizagem e reconhecimento. Alguns modelos para a percepção da fala em L2 propuseram que categorias de sons semelhantes, mas ligeiramente diferentes da L1, podem nunca ser totalmente adquiridas; eles tendem a constituir uma nova categoria intermediária no inventário de sons bilíngues que mescla as categorias semelhantes entre L1 e L2 (FLEGE, 1995; BEST, 1995). Sem dúvida, o reconhecimento de palavras faladas apresenta mais desafios em L2 do que em L1. Três desses desafios têm a ver com aspectos fonêmicos, silábicos e lexicais das representações bilíngues. Eles estão preocupados,

respectivamente, com (1) ativação de outros candidatos lexicais devido à coativação dos fonemas de L2, (2) interferência entre as representações fonológicas de L1 e de L2 devido a diferenças fonotáticas entre as duas línguas, e (3) diferenças nas características suprasegmentais de palavras de L1 e de L2. Vamos agora considerar cada um deles separadamente.

**(1) Representações fonêmicas bilíngues.** Se as representações de fonemas de um ouvinte não correspondem totalmente a todos os aspectos do sinal acústico, isso pode resultar na ativação de um *cohort* maior de candidatos a palavras, incluindo candidatos de L1 e de L2, em vez de apenas o idioma de destino. Essa superativação pode resultar em maior competição lexical (CUTLER, 2005). Por exemplo, idiomas como português, espanhol e italiano não fazem o contraste vocálico entre /ɪ/ e /i/ em seus inventários sonoros. Como consequência, as palavras em inglês “hit” e “heat” podem soar como homófonas para os falantes nativos dessas línguas (para vogais inglesas faladas por falantes nativos do BP, ver Nobre-Oliveira, 2007, e Rauber et al., 2005; por falantes nativos espanhóis, ver Morrison, 2006; por falantes nativos de italiano, ver Flege, McKay, & Meador, 1999). Assim, falantes bilíngues tenderão a ativar *cohorts* incluindo ambas as palavras em inglês durante a identificação de palavras, enquanto monolíngues em inglês ativariam apenas membros de *cohort* diretamente relacionados à palavra-alvo (ZWITSERLOOD, 1989). O aumento da competição lexical em L2 se tornaria evidente em latências de reconhecimento de palavras mais longas (e.g., WEBER & BROERSMA, 2012).

Uma consequência adicional dessa diferença na ativação do candidato lexical no domínio bilíngue é que os “falsos amigos” de L1 podem se tornar ativos. Se palavras com diferentes *inícios* entre L1 e L2 são percebidas como homófonas por bilíngues, não apenas falsos amigos lexicais do léxico L2 são ativados, mas palavras de L1 com sobreposição segmental “falsa” com palavras L2 também entram na competição, devido à natureza não seletiva de o léxico bilíngue (por exemplo, GOW & GORDON, 1995; ALLOPENNA, 1998; POST DA SILVEIRA, 2016). Assim, por exemplo, em bilíngues português-inglês, /pl/ pode ativar palavras de L1 em português, como *piquenique* e *pico* que contém a falsa sobreposição segmental /pi/ nas sílabas iniciais. Embora os ouvintes estivessem engajados no reconhecimento de fala de L2, a competição lexical com palavras de L1 resultaria na coativação dos competidores de L1 e de L2 e, também, seria afetada por seu significado. Em um contexto de frase, onde a semântica pode ajudar a desambiguar candidatos da L1, o grau de competição seria diminuído (FITZPATRICK & INDEFREY, 2010; WEBER & BROERSMA, 2012). Por exemplo, se o contexto do discurso estiver relacionado a *piquenique*, a ativação da palavra da L1 “piquenique” pode se tornar mais alta em relação a “pico”.

(2) Representações silábicas bilíngues. No processamento bilíngue, os padrões fonotáticos e os padrões de acento das palavras de L1 e de L2 interagem e interferem mutuamente. Ouvintes de L2 tendem a usar a silabificação e distribuições de padrão de acentuação de palavras de sua L1 na segmentação lexical de L2 (CUTLER, 2012). Línguas com ritmos sílaba e mora, como o português e o japonês, geralmente têm menos encontros consonantais do que as línguas com ritmo acelerado, como o holandês e o inglês (RASMUS & MEHLER, 1999). Falantes nativos dessas línguas percebem vogais epentéticas - que geralmente são as vogais mais reduzidas de seu inventário de sons - como situadas entre duas consoantes de encontros consonantais de L2 que não são encontrados em L1, formando uma nova sílaba CV átona (DUPOUX et al., 1999, 2011). Assim, esses falantes / ouvintes fazem correções perceptivas para estruturas fonotáticas que não existem em seu inventário de som nativo. Como outro exemplo, como nas plosivas do PB não são encontradas na posição silábica da coda, os brasileiros tendem a decodificar palavras como “big” e “topic” (estrutura fonotática CVC) como tendo um / i / epentético após a plosiva na posição da coda. Isso resulta na formação de duas sílabas CV, como /'bi.gi/ e /'to.pi.ki/, divergentes da sílaba CVC original da palavra em inglês. A segmentação fonotática baseada em L1 do sinal de fala de L2 pode resultar em segmentação lexical incorreta e / ou atrasada (CUTLER, 2012).

(3) Representações bilíngues de acentuação de palavras. Uma terceira diferença entre o reconhecimento de palavras faladas monolíngues e bilíngues tem a ver com os aspectos suprasegmentais da segmentação da fala. O acento na palavra é caracterizado por recursos suprasegmentais e segmentais. As categorias segmentais variam no número de unidades entre os segmentos (por exemplo, são 7 fonemas vocálicos discernidos no PB, enquanto em inglês são geralmente 14) e os padrões de acentuação da palavra têm diferentes distribuições de frequência lexical para diferentes idiomas. Como “o processamento da estrutura prosódica lexical pode ser explicado nos mesmos termos que o processamento de outras informações lexicalmente distintas” (CUTLER, 2012, p. 258), hipotetizamos que o acento de palavra de L2 pode resultar da fusão L1-L2 ou da dominância de L1 na atribuição de padrão de acento de palavra, semelhante ao que acontece com fonemas e padrões fonotáticos. Assim, pode haver uma mistura de estratégias perceptivas em L1 / L2 para avaliar o acento de palavra de L2 que vincula a evidência lexical no *input* aos padrões de acento da palavra na L1 e na L2 para aprendizagem e armazenamento. Se os ouvintes usarem representações compartilhadas da L1 e da L2 na compreensão e produção de palavras da L2, as pistas acústicas usadas no reconhecimento da acentuação da palavra podem ser semelhantes aos correlatos acústicos na produção.

Se os padrões de acentuação das palavras forem acusticamente semelhantes, mas não idênticos em duas línguas de um sistema bilíngue, os padrões de acentuação da

palavra da L2 podem ser representados em termos da categoria mais próxima da língua dominante, semelhante à categorização dos fonemas da L2 em termos de categorias da L1. Por exemplo, suponha que dois idiomas usem o pico da tonalidade (Pitch) para atribuir acento, mas em um idioma a inclinação é mais acentuada e no outro a inclinação é mais plana. No bilíngue, a apresentação real da inclinação como parte do acento da palavra da L2 pode agora ser mais plana em relação aos padrões de tom do monolíngue. Agora, suponha que haja congruência nos padrões. Por exemplo, se um padrão de acentuação de palavra é altamente frequente na L2 mais fraca e igualmente frequente na L1 dominante, é provável que o sistema bilíngue aceite esse padrão como o padrão de acentuação de palavra padrão nos dois idiomas e construirá uma categoria mesclada de L1-L2 (POST DA SILVEIRA ET AL., 2014). Portanto, se a L1 e a L2 têm a primeira sílaba das palavras como o padrão de acentuação da palavra mais frequente, como é o caso em inglês e holandês, o sistema bilíngue pode representar a acentuação da primeira sílaba como padrão para os dois idiomas (por exemplo, COOPER, CUTLER & WALES, 2002). Assim, o sistema dominante de atribuição de acentuação de palavra só se mostra quando o padrão de acentuação da primeira sílaba é produzido ou percebido de maneira imprecisa no reconhecimento ou produção de palavras (COLOMBO, 1992; COLOMBO & ZEVIN, 2009; POST DA SILVEIRA et al., 2014).

### 3. CONSIDERAÇÕES FINAIS

A discussão das seções anteriores sugere que o sistema cognitivo de L2 poderia ser organizado economicamente em relação ao armazenamento de detalhes sublexicais, como refinamento de categorias acústica, se contrastes que não são necessários para a inteligibilidade não forem ignorados pelos ouvintes. A representação, compreensão e produção bilíngue de características sublexicais da acentuação das palavras podem permanecer como as de L1, a menos que o ambiente linguístico exija que eles mudem por causa de mal-entendidos. Se a variabilidade no processamento sublexical pode ser facilmente corrigida via processamento *top-down*, ou seja, por meio da rota lexical, não há necessidade de uma especificação acústica precisa do componente sublexical da representação de acentuação da palavra da L2. Em contraste, no inglês nativo, a rota sublexical é usada proeminentemente para acessar o candidato lexical que melhor corresponde ao input acústico, porque a variabilidade das características acústicas é estreita e se encaixa nas categorias fonológicas do inglês para o acento de palavras em uma base *one-on-one*, com pouca necessidade de *feedback* via processamento *top-down* (MAGNUSSON et al., 2011; VAN HEUVEN, 2021).

Esse argumento sugere que os caminhos percorridos por bilíngues e monolíngues para acentuação das palavras e acesso lexical variam de acordo com a especialização de seu sistema. Monolíngues podem contar com as pistas subfonêmicas a nível sublexical para reconhecimento segmental e de acento de palavra para acessar rapidamente o candidato ideal de palavra; enquanto os bilíngues, porque suas representações acústicas de L2 para segmentos e acento de palavra constituem categorias acústicas mais amplas, confiam mais em *feedbacks* lexicais para decodificar o sinal de fala em fonemas e determinar o acento de palavra durante o reconhecimento de palavras de L2 (POST DA SILVEIRA et al., 2014; POST DA SILVEIRA, 2016, VAN HEUVEN, 2021).

Em suma, propomos que dois mecanismos centrais comuns são responsáveis pela representação e processamento do acento de palavra de L2 no reconhecimento e na produção de palavras: i) categorizações mescladas de categorias sublexicais segmentais de L1 e de L2 que se associam em representações específicas de acentuação de palavra de L2 e geram uma rota sublexical ruidosa para acesso lexical de L2; e ii) distribuições de frequência de padrão de acento de palavra mescladas entre L1 e L2 que geram representações de acento de palavra bilíngues divergindo de representações monolíngues, causando ruído no processo de acesso lexical de L2. Em trabalhos outros trabalhos (notadamente Post da Silveira et al., 2014; Post da Silveira, 2016, 2020a, 2021a) trouxemos informações sobre a pertinência dos mecanismos aqui expostos para produções de acento lexical em L1 e em L2 fundamentados com experimentos de produção em nomeação rápida de palavras e percepção com experimentos de eye-tracking que fazem uso do paradigma do mundo visual (ALOPENNA, MAGNUSON & TANENHAUS, 1998) com palavras impressas (Reinish, Jesse & MacQueen, 2010).

Ampliaremos a discussão sobre mecanismos de produção de acento lexical e acesso lexical em próximos trabalhos e, com a proposta desta série de trabalhos, pretendo tornar acessível a informação sobre modelos de percepção e produção lexical, dando ênfase a ao papel do acento lexical, para um número vasto de leitores que teriam acesso dificultado a esses conteúdos por não terem fluência de leitura em inglês, língua na qual a maioria dos textos sobre o tema estão escritos.

## REFERÊNCIAS

ALLOPENNA, P.D.; MAGNUSON, J.S.; TANENHAUS, M.K. Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of memory and language*, n. 38, v. 1, p. 419-439, 1998. DOI: <https://doi.org/10.1006/jmla.1997.2558>

BEST, C. T. A direct realist perspective on cross-language speech perception. In: STRANGE, W. (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Research*. York: Timonium. MD, 1995, p. 167-200.

COLOMBO, L. Lexical stress effect and its interaction with frequency in word pronunciation. *Journal of Experimental Psychology: Human Perception and Performance*, n. 18, v. 4, p. 987-1003, 1992. DOI: <https://doi.org/10.1037/0096-1523.18.4.987>

COLOMBO L, & ZEVIN, J. Stress Priming in Reading and the Selective Modulation of Lexical and Sub-Lexical Pathways. *PLoS ONE*, n. 9, v. 4, 2009. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0007219>

CHRISTOPHE, A.; MEHLER, J.; SEBASTIÁN-GALLÉS, N. Perception of prosodic boundary correlates by newborn infants. *Infancy*, n. 2, v.1, p.385-394, 2001.DOI: [10.1207/S15327078IN0203\\_6](https://doi.org/10.1207/S15327078IN0203_6)

CUTLER, A. Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*. 29. 201-220, 1986. Disponível em: <https://psycnet.apa.org/record/1988-22250-001>

CUTLER, A.; NORRIS, D. G. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, n. 14, v. 1, p. 113-121, 1988. DOI: <https://doi.org/10.1037/0096-1523.14.1.113>

CUTLER, A. *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press, 2012.

DAVIS, M. H. Lexical Segmentation in Spoken Word Recognition. Tese de Doutorado, Departamento de Psicologia, Birkbeck College, 2000. Disponível em: <https://www.mrc-cbu.cam.ac.uk/personal/matt.davis/thesis/davis.thesis.pdf>

DUPOUX, E.; KAKEHI, K.; HIROSE, Y.; PALLIER, C.; MEHLER, J. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology-human Perception and Performance*, n. 25, v. 6, p. 1568-1578, 1999. Disponível em: <https://hal.archives-ouvertes.fr/hal-02341221/document>

DUPOUX, E.; PARLATO, E.; FROTA, S.; HIROSE, Y.; PEPPERKAMP, S. Where do illusory vowels come from? *Journal of Memory and Language*, n. 64, v. 3, p. 199-210, 2011. DOI: <https://doi.org/10.1016/j.jml.2010.12.004>

ELMAN, J. L.; MCCLELLAND, J. L. Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, n. 27, v. 1, p. 143-165, 1988. Disponível em: <https://web.stanford.edu/~jlmcc/papers/ElmanMcC88.pdf>

FITZPATRICK, I.; INDEFREY, P. Lexical competition in non-native speech comprehension. *Journal of Cognitive Neuroscience*, n. 22, v.6, p. 1165-1178, 2010.

FLEGE, J. Second-language Speech Learning: Theory, Findings, and Problems. In: STRANGE, W. (Ed.). *Speech Perception and Linguistic Experience: Issues in Cross-language research*. Timonium, MD: York Press, 1995, p. 229-273.

FLEGE, J. E.; MACKAY, J. R. A.; MEADOR, D. Native Italian subjects' production and perception of English vowels. *Journal of the Acoustical Society of America*, n. 106, p. 2973-2987, 1999.

FRAUENFELDER, U.H.; SEGUI, J.; DIJKSTRA, A. Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance*, n.16, v.1, p.77-91, 1990. DOI: <https://doi.org/10.1037/0096-1523.16.1.77>

FRAUENFELDER, U. H.; PEETERS, G. Simulating the time course of spoken word recognition: An analysis of lexical competition in TRACE. In: GRAINGER, J.; JACOBS, A. M. (Eds.), *Localist connectionist approaches to human cognition*. Mahwah, NJ: Erlbaum, 1998, p. 101-146.

GOW, D.W.; GORDON, P.C. Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of experimental psychology: human perception and performance*, n. 21, v.1, p. 344-359, 1995. DOI: [10.1037/0096-1523.21.2.344](https://doi.org/10.1037/0096-1523.21.2.344)

HOLT, L. L.; LOTTO, A. J. Speech perception as categorization. *Attention, Perception & Psychophysics*, n. 72, v.1, p. 1218-1227, 2010. DOI: [10.3758/APP.72.5.1218](https://doi.org/10.3758/APP.72.5.1218)

JUSCZYK, P. W.; CUTLER, A.; REDANZ, N. Infants' reference for the predominant stress patterns of English words. *Child Development*, n. 64, v.1, p. 675-687, 1993. DOI: <https://doi.org/10.2307/1131210>

MAGNUSON, J.S.; MIRMAN, D.; LUTHRA, S.; STRAUSS, T.; Interaction in Spoken Word Recognition Models: Feedback Helps. *Frontiers in Psychology*, 2018. DOI: <https://doi.org/10.3389/fpsyg.2018.00369>

McCLELLAND, J. L.; ELMAN, J. L. The TRACE model of speech perception. *Cognitive Psychology*, n. 18, v.1, p. 1-86, 1986. DOI: [https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)

McCLELLAND, J. L.; RUMELHART, D. E. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, n. 88, p. 375-407, 1981.

McCLELLAND, J. L.; MIRMAN, D.; HOLT, L. L. Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, n. 10, v.8, p. 363-369, 2006. DOI: [10.1016/j.tics.2006.06.007](https://doi.org/10.1016/j.tics.2006.06.007)

MARSLÉN-WILSON, W. D.; WELSH, A. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, v.10, p.29-63,1978. DOI: [https://doi.org/10.1016/0010-0285\(78\)90018-X](https://doi.org/10.1016/0010-0285(78)90018-X)

MARSLÉN-WILSON, W. D.; TYLER, L. K. The temporal Structure of spoken language understanding. *Cognition*, v. 8, p. 1-71, 1980. DOI: [https://doi.org/10.1016/0010-0277\(80\)90015-3](https://doi.org/10.1016/0010-0277(80)90015-3)

MARSLÉN-WILSON, W. D. Activation, competition, and frequency in lexical access. In: ALTMANN, G. (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. Cambridge, Ma: MIT Press, 1990, p. 148-172.

MORRISON, G. S. L1-Spanish Speakers' Acquisition of the English /i /-/l/ Contrast: Duration-based perception is not the initial developmental stage. *Language and Speech*, n. 51, v. 4, p. 285-315, 2008. Disponível em: <https://journals.sagepub.com/doi/10.1177/002383090808099067>

NOBRE-OLIVEIRA, D. *The effects of training on the learning of American English vowels by native Brazilian Portuguese speakers*. Tese (Doutorado em Inglês), Universidade Federal de Santa Catarina, 2007.

NORRIS, D.; McQUEEN, J. M.; CUTLER, A. & BUTTERFIELD, S. The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, n.34, v.3, p. 19-243, 1997. DOI: [10.1006/cogp.1997.0671](https://doi.org/10.1006/cogp.1997.0671)

NORRIS, D. Shortlist: A connectionist model of continuous speech recognition. *Cognition*, n. 52, v.3, p. 189 -234, 1994. DOI: [https://doi.org/10.1016/0010-0277\(94\)90043-4](https://doi.org/10.1016/0010-0277(94)90043-4)

NORRIS, D.; & McQUEEN, J. M. Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, n. 115, v.2, p. 357-395, 2008. DOI: [10.1037/0033-295X.115.2.357](https://doi.org/10.1037/0033-295X.115.2.357)

PERRY, C., ZIEGLER, J. C. & ZORZI, M. Beyond single syllables: large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology*, 61, 106-151, 2010. DOI: [10.1016/j.cogpsych.2010.04.001](https://doi.org/10.1016/j.cogpsych.2010.04.001)

POST DA SILVEIRA, A., VAN HEUVEN, V., CASPERS, J., & SCHILLER, N.O. Dual activation of word stress from orthography: The effect of the cognate status of words on the production of L2 stress. *Dutch Journal of Applied Linguistic*, n. 3, v.2, p. 170-196, 2014. DOI: <https://doi.org/10.1016/j.dujal.3.2.05sil>

POST DA SILVEIRA, A. *Word stress in second language word recognition and production*. 1. ed. Enschede: Ipskamp, 2016. Disponível em: [https://www.researchgate.net/publication/312219967\\_Word\\_stress\\_in\\_second\\_language\\_word\\_recognition\\_and\\_production](https://www.researchgate.net/publication/312219967_Word_stress_in_second_language_word_recognition_and_production)

POST DA SILVEIRA, A.; SANDERS, E.; MENDONÇA, G.; DIJKSTRA, T. What Weighs for Word Stress? Big Data Mining and Analyses of Phonotactic Distributions in Brazilian Portuguese. In: VILLAVICENCIO, A. et al. (eds) *Computational Processing of the Portuguese Language*. PROPOR 2018. Lecture Notes in Computer Science. Springer, vol 11122, 2018, p. 399-408. DOI: [https://doi.org/10.1007/978-3-319-99722-3\\_40](https://doi.org/10.1007/978-3-319-99722-3_40)

POST DA SILVEIRA, AMANDA. Retrieving L2 word stress from orthography: Evidence from word naming and cross-modal priming. *Ilha do Desterro*, v. 73, p. 409-442, 2020. DOI: <https://doi.org/10.5007/2175-8026.2020v73n1p409>

POST DA SILVEIRA, A. "Representações lexicais e sublexicais do acento de palavra de L1 e de L2". In: VASCONCELOS, A.W.S. de (Org.). *Letras: Representações, Construções e Textualidades 2*. Ponta Grossa: Atena Editora, 2021, p. 1-10.

RASMUS, F.; MEHLER, J. Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America*, n. 105, v.1, p. 512-521, 1999. DOI: [10.1121/1.424522](https://doi.org/10.1121/1.424522)

RAUBER, A.; ESCUDERO, P.; BION, R. & BAPTISTA, B. O. The Interrelation between the Perception and Production of English Vowels by Native Speakers of Brazilian Portuguese. *Proceedings of Interspeech*, 2005, p. 2913-2916.

REINISCH, E., JESSE, A., & McQUEEN, J. M. Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *Quarterly Journal of Experimental Psychology*, 63 (4), p. 772-783, 2010. DOI: 10.1080/17470210903104412

RUMELHART, D. E.; J. L. MCCLELLAND. Interactive processing through spreading activation. In: PERFFETTI, C.; LESGOLD, A. (Eds.), *Interactive processes in reading*. Hillsdale NJ: Erlbaum, 1981, p. 37-60.

SEBASTIÁN-GALLÉS, N.; KROLL, J. Phonology in Bilingual Language Processing: Acquisition, Perception, and Production. In: MEYER, A.; SCHILLER, N. O. (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*. Berlin: Mouton de Gruyter, 2003, p. 279-318.

SULPIZIO, S., & McQUEEN, J. M. Italians use abstract knowledge about lexical stress during spoken word recognition. *Journal of Memory and Language*, 66, p. 177-193, 2012. DOI: <https://doi.org/10.1016/j.jml.2011.08.001>

VAN HEUVEN V.J.J.P. & VAN LEYDEN, K. Lexical stress and spoken word recognition, Dutch versus English. In: Dikken M. den, Cremers C. (Eds.) *Linguistics in the Netherlands 1996*. Amsterdam: John Benjamins. 159-170, 1996.

VAN HEUVEN, V. Resolving the prosody paradox. In: Post da Silveira, A.; Arantes, P. (Org.) *Prosódia e Bilinguismo vol. 1* (em editoração).

VROOMEN, J.; DE GELDER, B. Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, n. 21, v. 1, p. 98-108, 1995. DOI: <https://doi.org/10.1037/0096-1523.21.1.98>

WARREN, P.; MARSLÉN-WILSON, W. Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, n. 41, v.3, p. 262-75, 1987. Disponível em: <https://link.springer.com/article/10.3758/BF03208224>

WEBER, A.; BROERSMA, M. Spoken word recognition in second language acquisition. In: Chappelle C.A. (Ed.). *The Encyclopedia of Applied Linguistics*, Bognor Regis: Wiley-Blackwell, p. 5368-5375, 2012.

ZWITSERLOOD, P. The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, n. 32, v.1, 25-64, 1989. DOI: 10.1016/0010-0277(89)90013-9