DOI: http://dx.doi.org/10.31513/linguistica.2018.v14n3a21419
Recebido em: 01 de novembrol de 2018 | Aceito em: 29 de março de 2019

A APRENDIZAGEM DISTRIBUCIONAL NO PORTUGUÊS BRASILEIRO: UM ESTUDO COMPUTACIONAL

Pablo Picasso Feliciano de Faria[1]

Giulia Osaka Ohashi[2]

RESUMO

Neste artigo, tratamos do problema da aprendizagem de categorias de palavras durante o processo de aquisição. Nossa abordagem é computacional: construímos um modelo baseado em Redington et al. (1998) para investigar a informatividade da informação distribucional para a categorização de palavras. Os dados fornecidos ao aprendiz vêm de dois corpora de fala dirigida à criança em português brasileiro. Especificamente, os experimentos apresentados aqui avaliam a informatividade de várias janelas contextuais relativas a uma dada palavra-alvo, isto é, quais contextos são mais ou menos informativos sobre a categoria de uma palavra. Nossos resultados mostram que contextos locais são altamente informativos e que a informação distribucional é útil como fonte de informação categorial.

Palavras-chave: aquisição da linguagem; aprendizagem distribucional; categorias de palavras; modelagem computacional.

1 Doutor em Linguística pela Universidade Estadual de Campinas. E-mail: pablofaria@gmail.com.

2 Bacharel em Linguística, com ênfase em Linguística Computacional pela Universidade Estadual de Campinas. E-mail: giu.osaka@gmail.com.

ABSTRACT

In this paper, we address the problem of learning word categories during language acquisition. Our approach is computational: we built a model based on Redington et al. (1998) in order to investigate the informativeness of distributional information to the categorization of words. The data provided to the learner comes from two corpora of child-directed speech in Brazilian Portuguese. Specifically, the experiments presented here evaluate the informativeness of various contextual windows regarding a given target word, that is, which contexts are more or less informative of a word category. Our results show that local contexts are highly informative and that distributional information is useful as a source of categorial information.

Keywords: language acquisition; distributional learning; word categories; computational modeling.

1. Introdução

Várias questões movem a área de aquisição da linguagem, com vistas a compreender como se dá o processo de aquisição das línguas humanas por crianças. Por exemplo, o que explica a aquisição rápida, espontânea e uniforme (quanto ao desenvolvimento e ao conhecimento linguístico adquirido), como a que se verifica em qualquer criança típica de qualquer comunidade falante no mundo? Quais estratégias de aprendizagem estão disponíveis à criança para que esta tenha sucesso? Neste trabalho, apresentamos resultados experimentais que iluminam alguns aspectos dessas questões, relativos ao processo de categorização de palavras a partir de informação distribucional, tais como outros trabalhos encontrados na literatura (Redington *et al.,* 1998; Mintz *et al.*, 2002).

Redington *et al.* (1998) foram os primeiros a conduzir um estudo computacional para modelar e investigar o quanto de informação sobre categorias de palavras a criança poderia extrair apenas monitorando a distribuição das palavras nos enunciados, sem olhar para aspectos morfológicos e semânticos, entre outros com possível papel nesse aprendizado. Além de ser um estudo pioneiro, este é ainda o estudo mais abrangente encontrado na literatura até o presente momento. Por esta razão, neste trabalho partimos de Redington *et al.* (1998) e investigamos via modelagem computacional o potencial da informação distribucional, em diferentes condições experimentais, para categorização de palavras em classes lexicais no português brasileiro (doravante, PB).

No que segue, apresentamos resultados parciais de estudos em andamento, no caso, avaliações

dos contextos distribucionais mais informativos à criança. Nossos estudos visam não apenas replicar, para o português, modelagens como a de Redington *et al*. e outros (como o de Mintz *et al.*, 2002, por exemplo), mas ainda investigar questões não cobertas em tais estudos. Os dados utilizados nas simulações foram retirados da base *CHILDES* (MacWhinney, 1989) e da Coleção "Projeto Aquisição da Linguagem Oral"[3]. Ao se propor a analisar dados do PB, este trabalho contribui para uma avaliação translinguística de achados (para o inglês) em Redington *et al*. (op.cit.). Nossos resultados contribuem assim para o avanço das teorias de aquisição e, assim como no estudo original, indicam que a informação distribucional é útil na aprendizagem das categorias (sintático-)lexicais do PB.

Este artigo está organizado da seguinte forma: na **seção 2**, é feita uma introdução breve sobre aspectos teóricos da aquisição da linguagem envolvendo o "Argumento da Pobreza de Estímulos" e como isso se relaciona com a questão da aprendizagem distribucional. Nesta seção, ainda introduzimos a abordagem computacional. Na **seção 3**, que trata de materiais e métodos, são apresentadas informações sobre o corpus utilizado neste estudo e é feita uma apresentação detalhada do método distribucional implementado. Na **seção 4**, apresentamos os resultados e a discussão dos mesmos. Finamente, a **seção 5** traz considerações finais.

2. A aquisição da linguagem, dotação inata e o papel da experiência

A aquisição da linguagem é parte do processo natural de desenvolvimento humano, isto é, todas as crianças típicas expostas à língua em uma comunidade falante irão desenvolver a língua fluentemente em um período de tempo similar. Mesmo em alguns casos de comprometimento cognitivo, como no caso da Síndrome de Williams, crianças ainda se mostram capazes de adquirir a língua, o que tem sido tomado como indício de uma dissociação entre a linguagem e outros domínios cognitivos (Rossi *et al.*, 2006).

Uma das hipóteses para explicar esse fenômeno único na cognição humana – dado que em nenhum outro domínio cognitivo observa-se um processo dessa natureza – é a da existência de uma Gramática Universal (GU) (Chomsky, 1986), que seria parte da dotação inata da espécie humana, responsável por "guiar" o processo de aquisição na medida em que informaria a criança sobre o que é uma língua natural possível e sobre como explorar os dados da experiência para adquirir sua língua nativa. Chomsky e outros adeptos da Gramática Gerativa afirmam ainda que as línguas naturais são

3 Inventário do Projeto Aquisição da linguagem Oral. Org. Vania Regina Personeni. Campinas: CEDAE/ IEL, s.d. 33p.; Plataforma de Documentos Sonoros do Cedae. Disponível em: <https://goo.gl/Jop0NU>. Acesso em: 15/10/2018.

específicas à espécie humana e que a aquisição apresenta características próprias do desenvolvimento biológico, tais como o desenvolvimento espontâneo, sequencial, uniforme e exitoso entre as crianças, apesar da variabilidade da experiência. Esta visão uniforme da aquisição e do conhecimento linguístico não é, vale ressaltar, um ponto pacífico em toda a área e tem sido crescentemente questionado na literatura, como se vê, por exemplo, em Evans & Levinson (2009).[4] Como argumentamos mais a frente, porém, a resolução desse debate depende, em parte, de estudos como o aqui proposto.

Outro argumento importante para a hipótese de uma GU inata é o conhecido "argumento da pobreza de estímulos" (revisado recentemente em Berwick et al., 2011). Segundo este, a experiência linguística da criança compõe-se de dados incompletos e degradados (contendo reformulações, interrupções, barulhos etc.), aleatórios (isto é, sem sistematização) e, ainda, sem evidência negativa significativa (isto é, ensino ou correção explícitos da gramática da criança por parte dos falantes proficientes). Além disso, a criança em fase de desenvolvimento não está concentrada somente em aprender a falar, mas também em comer, aprender a andar etc. Assim, haveria uma enorme lacuna entre a qualidade da experiência linguística da criança e a complexidade do conhecimento gramatical adquirido, lacuna esta intransponível com base em procedimentos indutivos gerais. Muitos estudiosos defendem, portanto, que a aquisição da linguagem seria praticamente impossível sem uma dotação inata específica.

Um dos grandes empecilhos ao avanço dessa discussão sobre "natureza x nutrição" é sair do plano especulativo para determinar com precisão qual a contribuição de cada fator. E talvez a opção mais viável inicialmente, dada a dificuldade de compreender as propriedades e processos cerebrais, seja determinar com precisão o que a nutrição fornece, isto é, quais (e o quanto de) informações sobre a linguagem estão disponíveis e são acessíveis ao aprendiz da língua. É nesta frente que o estudo apresentado neste trabalho se insere, na medida em que tenta mensurar o grau de informatividade da informação distribucional sobre classes de palavras que pode ser extraída dos enunciados que a criança ouve.

2.1. A utilidade da informação distribucional

Visto que o presente estudo teve como objetivo principal replicar experimentos apresentados em Redington *et al*. (1998), nesta seção sumarizamos o método distribucional proposto pelos autores. Os autores citam Harris (1954), o qual introduz o conceito de "distribuição" da seguinte forma: "a

4 Agradecemos a um dos pareceristas por chamar a atenção para este ponto.

distribuição de um elemento será entendida como a soma de todos os contextos em que ocorre" (p. 146). Harris lista quatro indícios de uma estrutura distribucional na linguagem: (i) a ocorrência dos elementos nos enunciados é determinada uns pelos outros; (ii) tais restrições se mantém para todas as ocorrências dos itens (i.e., são de fato estruturais); (iii) pode-se estabelecer (probabilisticamente) a ocorrência de um elemento qualquer em relação a outro elemento, até o limite de exatidão determinado pela relação entre suas classes; e (iv) as restrições sobre a ocorrência relativa de cada elemento são melhor descritas como base em suas classes do que como medidas simples, globais e separadas para cada um, numa forma de enumeração direta.

Harris (op.cit.) defende que a estrutura distribucional existe de fato na linguagem e existiria também *nos falantes*. Ainda segundo Harris, é esperado o comportamento dos falantes indique sua percepção da estrutura distribucional. O processamento de enunciados (na compreensão ou na produção) seria, portanto, baseada em relações distribucionais. A esse respeito, muitos estudos sobre aquisição da linguagem demonstram o papel da informação distribucional na aquisição de palavras (Brown, 1957; Landau & Gleitman, 1985; Naigles, 1990; Hohle *et al.*, 2004; Bernal *et al*., 2007, entre outros).

A este respeito, Redington *et al*. (1998) comentam duas críticas comuns à hipótese da utilidade da informação distribucional. A primeira é a de que a utilidade dessa informação seria óbvia, dado que categorias sintáticas são determinadas por sua distribuição. Esse argumento é incorreto, no entanto, pois não reconhece a diferença entre a natureza das informações distribucionais usadas pelos linguistas e a informação distribucional que estaria disponível para a criança: no primeiro caso, informações podem ser elicitadas de modo preciso, enquanto crianças tem à disposição dados aleatórios, parciais e afetados por ruídos (dados degenerados). A tarefa da criança é imensamente mais complexa, daí uma segunda crítica na direção oposta: a informação distribucional não é útil para aprender categorias sintáticas.

Um ataque influente deste tipo é o de Pinker (1984), em que o autor afirma que (i) a quantidade de relações distribucionais possíveis a considerar estaria fora do alcance de mecanismos de aprendizagem, (ii) que muitas propriedades superficiais são irrelevantes, (iii) que mesmo dentre as propriedades relevantes, línguas variam muito com relação a quais mobilizam, e (iv) que correlações locais "espúrias" emergem em dados como "João come maçã" e "João come lentamente" (adaptação nossa), em que o aprendiz concluiria que "maçã" e "lentamente" são da mesma categoria. Porém, como argumentam Redington *et al.* (op.cit.), nenhum destes argumentos convecem, pois (i) não é preciso assumir que o aprendiz busque cegamente por qualquer propriedade possível, (ii) o fato de

haver propriedades irrelevantes não impede que se aprenda com as que são, (iii) a variação entre línguas também não pode ser obstáculo a este tipo de estudo, pelo contrário, o torna essencial, e (iv) cabe aos estudos mostrar que aprendiz pode superar tais problemas locais a partir de mecanismo psicologicamente plausíveis.

Portanto, determinar a utilidade da informação distribucional é uma questão empírica, o que justifica estudos como o de Redington *et al.* (1999) e outros, como o apresentado aqui. Dado que este é um problema particularmente tratável computacionalmente e que a cada dia são disponibilizados mais e mais corpora de aquisição da linguagem, sua investigação se torna muito atrativa e benéfica para o desenvolvimento da área e da compreensão destas questões.

2.2. Modelos computacionais como meio de investigação

Um meio cada vez mais utilizado para investigação de questões de aquisição da linguagem é o desenvolvimento de modelos computacionais. Tais modelos computacionais podem ser vistos como aproximações dos processos psicolinguísticos que se dão em crianças no processo de aquisição da linguagem (Kaplan *et al.*, 2008; Pearl, 2010; Yang, 2011). Na modelagem computacional (Marr, 1982 *apud* Pearl, 2010), é necessário lidar com três níveis de processamento de informação, sendo os dois primeiros responsáveis por tratar de questões psicolinguísticas, enquanto o terceiro trata da "engenharia" da construção do modelo, isto é, de como implementar o modelo computacionalmente.

O primeiro nível é o computacional e diz respeito à descrição formal do problema a ser modelado. Dessa forma, é o nível que dialoga mais fortemente com as teorias linguísticas, de aquisição e psicolinguística. O segundo nível é o algorítmico, em que as condições e meios necessários para a solução do problema de aprendizagem são especificados, isto é, os procedimentos de aquisição que operam sobre os dados de entrada e com base na "dotação" assumida para o aprendiz. Este nível se relaciona, assim, com a *teoria de aprendibilidade*, na medida em que precisa dar respostas claras e objetivas para as seguintes questões:

a) O que é aprendido, exatamente?

b) Quais tipos de hipóteses o aprendiz é capaz de entreter?

c) Como os dados da língua-alvo são apresentados ao aprendiz?

d) Quais restrições governam o modo como o aprendiz atualiza suas conjecturas em resposta aos dados?

e) Sob quais condições, exatamente, dizemos que o aprendiz obteve sucesso na tarefa de aprendizagem da linguagem?

Responder a estas questões implica, para além de explicitar formalmente as assunções, tornar o modelo capaz de refletir aspectos teóricos e observações empíricas do processo de aquisição, além de fazer novas previsões sobre o processo. Assim, uma das principais virtudes dos modelos computacionais é esta necessidade de explicitude quanto a propriedades e mecanismos propostos ou assumidos, aspectos normalmente tratados de modo relativamente informal na teoria.

Finalmente, para que um modelo seja considerado plausível e substancial, é necessário almejar atender a algumas condições pertinentes, tais como os critérios apontados por Pinker (1979), a saber:

i. *Aprendibilidade*: o aprendiz deve aprender o que é esperado;

ii. *Equipotencialidade*: deve (potencialmente) se aplicar a outras línguas;

iii. *Entrada*: deve fazer assunções plausíveis sobre os dados de entrada;

iv. *Tempo*: deve aprender no mesmo tempo ou com base em uma quantidade de dados equivalente ao que uma criança típica dispõe;

v. *Desenvolvimental*: deve exibir um percurso similar, incluindo progresso e desvios, ao de uma criança;

vi. *Cognitiva*: deve assumir recursos cognitivos equivalentes ao que uma criança tem à sua disposição.

Não necessariamente um modelo irá responder satisfatoriamente a todos estes critérios, não apenas pelos desafios intrínsecos, mas também por lacunas teóricas relativas aos vários aspectos envolvidos. No caso do estudo apresentado aqui, que trata de como a criança aprende as categorias léxico-sintáticas das palavras, dentre os critérios supracitados, os que mais diretamente se aplicam são os critérios (iii), (vi) e, especialmente, o (ii), visto que aqui se replica, para o português brasileiro, experimentos feitos em Redington *et al*. (1998) originalmente para o inglês.

3. Materiais e Métodos

Nosso estudo computacional foi desenvolvido em linguagem *Python* e faz uso de diversas bibliotecas, tais como a *Natural Language Toolkit* (NLTK), com diversas funcionalidades para processamento de linguagem natural, e a *Numpy*, a *SciPy* e a *Math*, que são úteis para as análises estatísticas e funções de agrupamento hierárquico, necessárias para a implementação do método distribucional. Para a realização das simulações foi necessário também compilar um corpus de dados do PB contendo fala dirigida à criança (corpus FDC), o qual descrevemos a seguir.

3.1. Tratamento do corpus

A compilação do corpus utilizado nas simulações envolveu, primeiramente, a obtenção dos dados nas suas respectivas fontes. No caso da base CHILDES, os arquivos podem ser baixados diretamente do site respectivo, em formato texto. Para os dados do "Projeto Aquisição da Linguagem Oral", foi necessário realizar diversas visitas ao CEDAE/IEL/Unicamp, para obtenção de todo o corpus transcrito em arquivos em formato PDF, a partir dos quais geramos versões em formato texto.

A segunda etapa de preparação consistiu em preparar os dados para processamento. Isso incluiu remover metadados, normalizar identificações de falantes, regularizar a estrutura dos enunciados de cada falante e excluir a fala das crianças, além de remover marcações diversas, comentários e observações contextuais dos investigadores inseridas nas transcrições. Além disso, principalmente para o corpus do CEDAE, foi necessário padronizar a ortografia, que ora refletiam características da fala, ora padrões ortográficos normativos. Isso foi feito automaticamente e parcialmente, de modo a cobrir os casos mais recorrentes, pois uma padronização completa demandaria uma ampla revisão das transcrições face às gravações sonoras.

Além dos dados de fala, era necessário compilar também uma categorização de referência das palavras-alvo, para que a performance do método distribucional pudesse ser avaliada. Para isso, usamos os dados etiquetados (i.e., anotados morfossintaticamente) do *Corpus Histórico do Português Tycho Brahe* (doravante, CTB), constituído por versões eletrônicas anotadas de textos em português escritos por autores nascidos entre 1380 e 1881.[5] As palavras-alvo não contempladas pelo CTB necessárias aos experimentos foram classificadas manualmente de acordo com o seu papel sintático mais comum. Esses casos eram majoritariamente não ambíguos, consistindo de palavras da contemporaneidade, tais

5 Disponível em http://www.tycho.iel.unicamp.br/~tycho/corpus/texts/pos.zip. Acessado em junho/2017.

como "computador", "televisão" etc., substantivos próprios ("Raquel" e "Fernando", por exemplo) e formas diminutivas ("menininho" e outras). Pseudo-palavras como "popó", "blim" e "cocó" foram deixadas sem classificação. Nesse sentido, optamos por procedimentos análogos aos de Redington *et al*. (op.cit.). Ao final do pré-processamento, o corpus utilizado consistiu de 1,43 milhão de pares palavra/etiqueta.

No entanto, lidar com o PB colocou problemas metodológicos e conceituais não enfrentados em Redington *et al*. (op.cit.), uma vez que a morfologia do PB se mostra mais diversa e rica que a do inglês, expressando através de morfemas relações predicativas que demandam expressões multipalavras no inglês, como em "menininho" *vs.* "little boy" (daí a importância de estudos translinguísticos). Em particular, casos de flexões dos substantivos no diminutivo, no aumentativo e por gênero não foram considerados como uma mesma palavra – uma simplificação aparentemente plausível. Em primeiro lugar, por ser possível que formas flexionadas se especializem com sentidos distintos, como "calça" e "calcinha", diferentemente do par "cachorro" e "cachorrinho", por exemplo.[6] Em segundo lugar, pelo fato de que o conhecimento da morfologia flexional é uma aquisição mais tardia da criança e o modelo não cobre (a aquisição da) análise morfológica das palavras. Todos os itens flexionados foram, portanto, adicionados manualmente à categorização de referência.

Finalmente, fizemos a normalização da pontuação seguindo o procedimento em Redington *et al*. (op.cit.): removeu-se toda a pontuação intermediária dos enunciados e toda pontuação final (onde ocorria de fato) foi transformada em um simples ponto-final. Ao final do processo, obtivemos o *corpus FDC*, juntando dados do CEDAE e da base CHILDES, num total aproximado de 1,4 milhão de tokens (incluindo pontuação final).[7]

3.2. O método distribucional

Com o intuito de demonstrar que as propriedades distribucionais das palavras podem ser altamente informativas, no que diz respeito à categoria sintática, e como essa informação pode ser

6 Como apontado por um dos pareceristas, há ainda mais complicações aí, tais como o uso do diminutivo para produzir efeitos metafóricos, como em "ela é um cachorrinho do patrão", dando o sentido de submissão.

7 Em termos comparativos ao estudo de Redington *et al*., nosso corpus do PB equivale a pouco mais de 50% do corpus utilizado ali (1,4 milhões de palavras contra 2,5 milhões do estudo original). O ideal seria compilarmos um corpus de tamanho similar, algo que temos como objetivo. De todo modo, mesmo com um corpus similar, a comparação seria ainda aproximada, algo inevitável quando se trata de reimplementações de métodos e, no caso, de um corpus de outra língua.

extraída de modo mecanismos psicologicamente plausíveis, Redington *et al.* (1998) propõem três estágios para desenvolver esse tipo de análise: (i) medir os contextos de distribuição em que cada palavra ocorre; (ii) comparar o contexto de distribuição para pares de palavras; e (iii) agrupar palavras com distribuições de contextos similares.

O primeiro estágio envolve coletar o contexto de ocorrência das palavras-alvo, a saber, estatísticas de co-ocorrência entre uma dada palavra-alvo e palavras em seu entorno, armazenando estes dados estatísticos em *tabelas de contingência*. Nestas, cada linha representa a distribuição de uma palavra e cada coluna registra a quantidade de vezes que cada palavra de contexto aparece numa dada relação distribucional (p.e., imediatamente antecedente) com cada palavra-alvo. Se mais de uma relação distribucional é avaliada para cada palavra-alvo (p.e., as palavras imediatamente precedente e imediatamente sucessora), então, mais tabelas de contingência são construídas e as linhas de uma dada palavra-alvo em cada tabela são concatenadas, formando uma representação vetorial da distribuição observada, chamado de *vetor de contexto*.

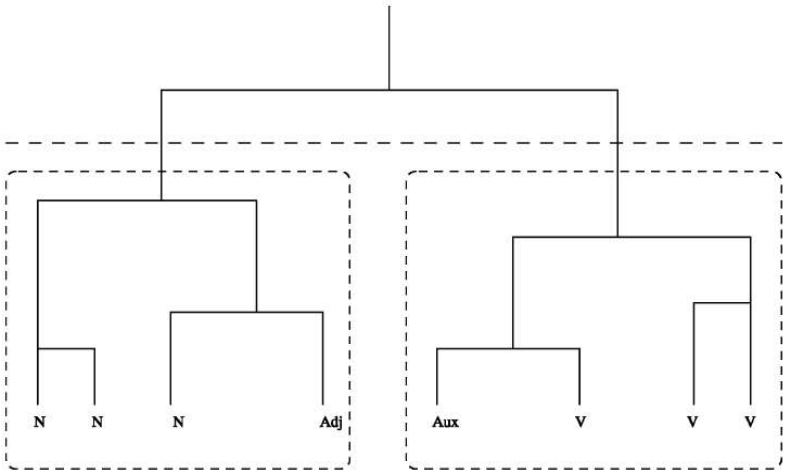
O segundo estágio envolve avaliar a similaridade entre os vetores de contexto e, portanto, entre as palavras-alvo. Segundo Redington *et al.* (1998), o vetor de contexto de cada palavra-alvo pode ser pensado como um ponto em um espaço multidimensional de possíveis distribuições de contextos. Assim, é possível esperar que palavras da mesma categoria sintática tenham distribuições similares, isto é, ocupem posições próximas nesse espaço. Para medir a similaridade entre duas palavras quaisquer, os autores utilizaram o *coeficiente de correlação de postos de Spearman (ρ)*[8] aplicado sobre seus vetores de contexto respectivos. A partir dessas medidas, as palavras podem então ser agrupadas, o que representa o terceiro estágio.

Nesta última etapa, é utilizada a *análise de cluster hierárquica padrão* (Sokal & Sneath, 1963 apud Redington *et al.*, op.cit.), conhecida como *cluster de link médio*. O algoritmo começa combinando os dois itens que estão mais próximos de acordo com a métrica de similaridade. Uma vez formado o primeiro *cluster*, o algoritmo segue buscando pelos itens com maior similaridade para agrupá-los, podendo ser tanto outras palavras-alvo, como também *clusters* entre si, ou combinações de *clusters* com palavras-alvo. A distância entre dois *clusters* quaisquer é a média das distâncias entre os membros de cada um. O algoritmo termina quando um *cluster* final é obtido, que inclua todos os demais. Ao final, forma-se uma estrutura hierárquica que estabelece agrupamos para diferentes níveis de similaridade e que pode ser representada como um *dendrograma*, como exemplificado na Figura

8 Há várias medidas possíveis para a similaridade, não sendo esta a única. Para mais detalhes, ver Redington *et al.* (1998, pgs. 436-438).

1. A depender do ponto (nível de similaridade) onde é recortado o dendrograma, obtém-se um número distinto de agrupamentos (na Figura 1, a linha horizontal tracejada seleciona dois grandes grupos).

Figura 1 - Dendrograma de agrupamentos extraído de Mintz et al. (2002, p. 400).



3.2.1. Medidas de performance do método

É preciso avaliar a performance dos métodos através de medidas objetivas. Duas medidas comuns utilizadas são as de *precisão* e *completude*, muito comuns nesse tipo de estudo. A primeira mede a proporção de pares de palavras colocadas no mesmo grupo pelo método que de fato são da mesma categoria, segundo a classificação de referência. Ou seja, se o método propôs 50 pares de palavras, mas apenas 10 são corretos, sua precisão seria de 20%. A segunda mede a quantos pares de palavras que deveriam estar no mesmo grupo que o método acertou. Em outras palavras, se havia 40 pares corretos a conjecturar, mas apenas 10 foram conjecturados, então a cobertura é de 25%. Estas duas medidas são complementares e, normalmente, quando se aumenta a precisão, perde-se cobertura e vice-versa.

Por esta razão, é preciso uma medida integrada, que balanceie estas duas. Redington *et al*. (1998) propõem uma medida que chamam de *informatividade*, enquanto Mintz *et al.* (2002) propõem uma medida que chamam de *pureza*, ambas com suas próprias justificativas. No presente estudo, usamos a medida *F*, conhecida como média harmônica da precisão e da completude, dada pela fórmula geral , em que optamos por favorecer a precisão sobre a completude, usando um coeficiente de . Essa escolha se deve ao fato de que as categorias sintáticas apresentam uma peculiaridade: as lexicais, como nomes e verbos, são categorias potencialmente ilimitadas em número de elementos, enquanto as "funcionais" (ou gramaticais) são classes "fechadas", isto é, tem um conjunto finito de elementos

(ver Tabela 1). Isso produz um efeito: se o método acerta muitos dos itens das categorias lexicais, ele terá uma ótima cobertura, porém ao custo da precisão, visto que será muito pouco preciso para as categorias funcionais. Por isso, ao favorecer a precisão através do coeficiente, tentamos balancear melhor essa relação.

Nos experimentos apresentados mais à frente, a performance é calculada para vários níveis de similaridade, variando de 0 a 1, com intervalos de 0,01 (ou seja, 100 recortes). Para cada um, são calculadas as três medidas indicadas acima. A partir destas, identifica-se o nível de similaridade que maximiza a performance, em termos da medida *F*.

3.2.2. A classificação de referência

Para avaliar a performance do método distribucional é preciso ter uma classificação de referência contra a qual a classificação obtida pelo método possa ser contrastada. Embora muitas palavras possam ter mais de uma categoria, nesta modelagem o problema se restringe a estabelecer a mais provável para uma dada palavra. Sem dúvida, aprender a distinguir entre as instâncias de uma palavra que pertencem a categorias distintas é uma das tarefas da criança, durante a aquisição. Nesta modelagem, porém, a classificação de referência estabelece a categoria mais provável para cada palavra-alvo, assumindo que esta informação seja relativamente estável entre amostras da língua grandes o suficiente.[9]

Assim, a partir de um corpus contendo 1,43 milhão de palavras[10], são classificadas as palavras-alvo para cada um dos experimentos descritos na seção seguinte. De modo a usar um sistema de anotação mais básico, como o adotado em Redington *et al.* (op.cit.), foi necessário fazer uma conversão do sistema de categorias do CTB, como é possível observar na Tabela 1. Para o conjunto das 1000 palavras mais frequentes do corpus FDC usadas como palavras-alvo nos experimentos, são exibidas as categorias, as etiquetas (base) correspondentes no CTB, algumas palavras de exemplo e o número de palavras naquela categoria. Note que para as categorias envolvendo "contrações", para as quais não fica claro a quê se referem, não houve conversão, não sendo, portanto, utilizadas no estudo

9 Nos parece que a classificação ideal seria uma feita sobre o próprio corpus FDC utilizado na aprendizagem, visto que esta fala pode apresentar características lexicais particulares. Porém no momento não dispomos dessa anotação.

10 Portanto, um corpus com cerca de 10% do tamanho do corpus utilizado em Redington *et al.* (1998). Em estudos subsequentes, na ausência de um padrão "ouro" para o próprio corpus FDC (ver nota de rodapé anterior), será importante ampliar este corpus de referência.

apresentado aqui. Para a conversão de etiquetas compostas do CTB, como P+D ou VB+CL, optamos por utilizar a primeira etiqueta (nos exemplos, P e VB, respectivamente) para fins de conversão.

Tabela 1 - Categorias, exemplos e quantidades para o conjunto das 1000 palavras mais frequentes do corpus FDC.

Categoria	Etiquetas do CTB	Exemplo	n
Substantivo	N, NPR	ademir, adriana, ajuda	375
Adjetivo	ADJ, OUTRO	alto, amarelo, baixo	82
Numeral	NUM	cinco, dez, duas	14
Verbo	VB, HV, ET, TR, SR	abre, abrir, abriu	331
Artigo	D	a, aquele, os	45
Pronome	CL, SE, DEM, PRO, PRO$, SENAO, QUE, WADV, WPRO, WD, WPRO$, WQ	aonde, aquilo, cadê	53
Advérbio	ADV, Q, NEG, FP	agora, ainda, algum	62
Preposição	P	até, co, com	11
Conjunção	CONJ, CONJS, C	como, e, enquanto	11
Interjeição	INTJ	ah, ahn, ai	16
Contração simples	-		
Contração complexa	-		

3.2.3. Um "piso" classificatório

Para demonstrar que a classificação produzida pelo método realmente produz informação potencialmente útil ao aprendiz, não apenas precisamos medir sua performance contra uma classificação de referência, como explicado anteriormente, mas é também necessário mostrar que ela produz classificações melhores do que uma classificação produzida aleatoriamente. Para isso, seguimos o mesmo procedimento indicado em Redington *et al.* (1998): para cada nível de similaridade avaliado, mantém-se fixo o número de agrupamentos obtidos, mas as palavras são *aleatoriamente* redistribuídas entre os grupos e a performance é recalculada. Isso é feito dez vezes para cada nível de similaridade e a performance final considerada é aquela obtida pela média das dez classificações aleatórias. Isso significa que para cada experimento, a piso classificatório precisa ser recalculado.

4. Resultados e discussão

Nesta seção, iniciamos com uma análise qualitativa do experimento "padrão", a partir da qual discutimos a alcance geral do método. Em seguida, apresentamos os resultados quantitativos da manipulação da janela de contexto assumida, que mostra o grau de contribuição de outras palavras de contexto, quando em certas posições no entorno das palavras-alvo.

4.1. Análise qualitativa

No experimento padrão, são utilizadas as 1000 palavras mais frequentes como palavras-alvo a serem classificadas. Como palavras de contexto, tomam-se as 150 palavras mais frequentes. A janela de contexto considerada inclui as duas palavras imediatamente precedentes à palavra-alvo e também as duas imediatamente posteriores, tomando-se todo do corpus FDC e suas 1,15 milhão de palavras. Como são 4 posições contextuais a considerar e 150 palavras de contexto possíveis, os vetores de contexto das palavras-alvo terão, cada um, 600 elementos. Cada elemento corresponde à frequência de uma dada palavra de contexto numa dada posição contextual. Para este experimento, toda a pontuação final é removida e os enunciados são concatenados, de modo que o corpus é tratado como uma única longa sentença. Na Figura 2 vemos 12 dos 17 agrupamentos obtidos no experimento padrão, para o nível de similaridade 0,39, no qual obtém-se uma medida $F = 0{,}67$ (precisão de 0,73 e completude de 0,35), muito acima do piso classificatório, como pode ser visto no Gráfico 1.

Gráfico 1 - Resultados quantitativos do experimento padrão.

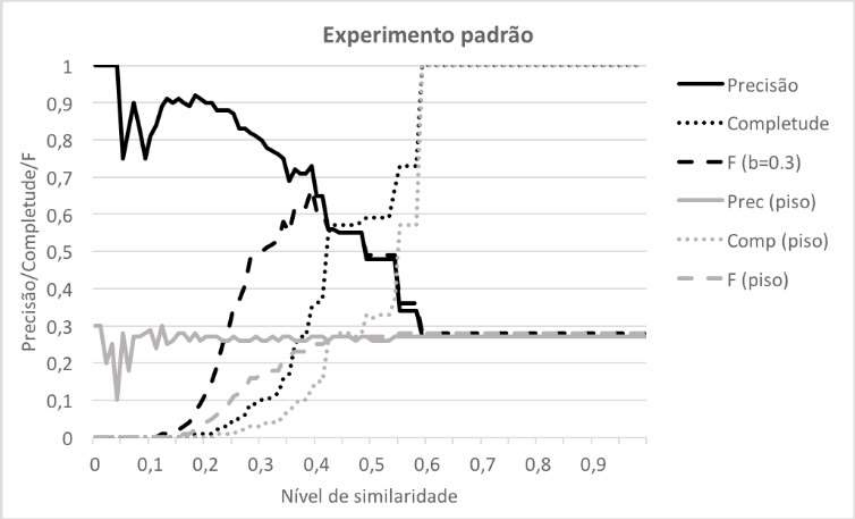
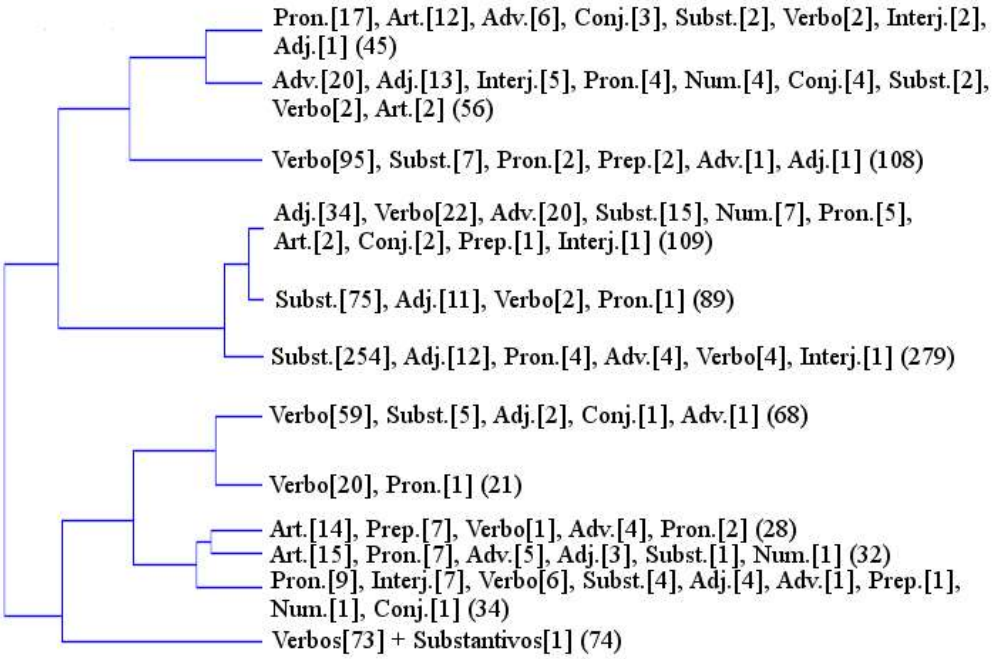


Figura 2 - Os agrupamentos no nível 0,39 de similaridade para o corpus FDC. Os agrupamentos foram nomeados manualmente com as categorias de referência de seus itens, incluindo suas quantidades. Apenas agrupamentos com 20 ou mais elementos são exibidos (12 de 17).



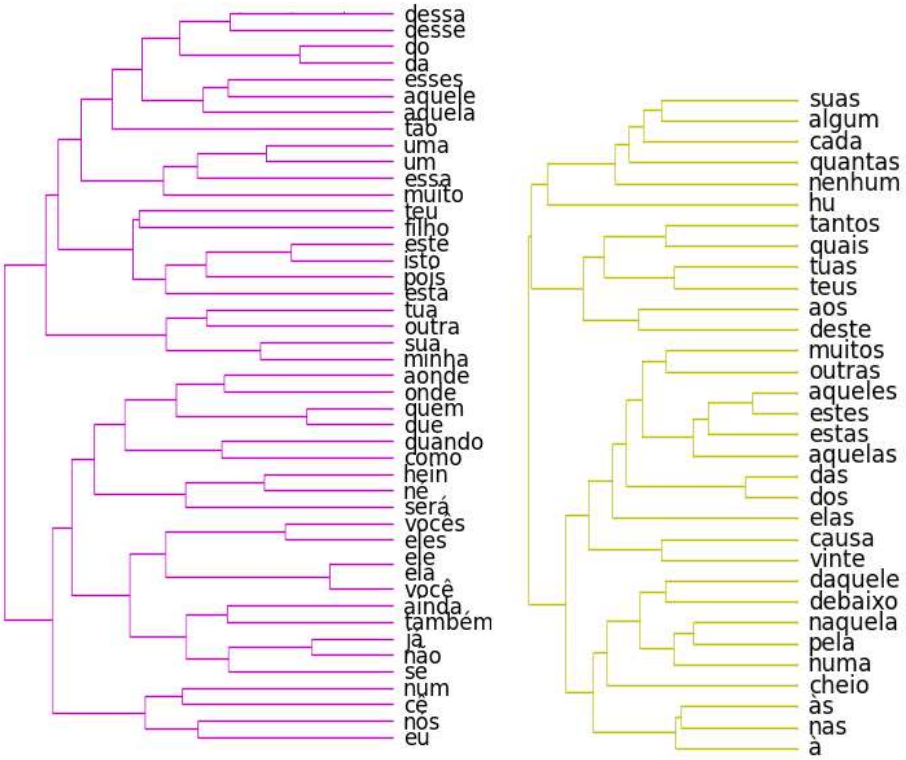
Exibimos apenas os agrupamentos com 20 ou mais elementos. Uma primeira impressão é a de que os grupos são heterogêneos. Porém, ao observar com calma, vemos alguns grupos mais "puros" que outros. Dois ressaltam: o grupo de 21 elementos, sendo 20 verbos e 1 pronome, e o grupo de 74 elementos, sendo 73 verbos e 1 substantivo. No caso do primeiro, que consiste de verbos flexionados, especialmente os verbos "ser", "estar", "fazer" e "ter", aparece como "intruso" o pronome "lhe". A única explicação que nos vem à mente é a de que tal resultado se deve a limitações do corpus, isto é, dados insuficientes para capturar corretamente a categoria de "lhe"[11]. Retomaremos o segundo agrupamento mais adiante.

Para grupos envolvendo mais as classes funcionais, vemos de fato uma grande heterogeneidade, embora ainda assim seja possível ver tendências. Na parte debaixo da Figura 2, vemos 3 agrupamentos próximos, envolvendo primordialmente artigos, pronomes e preposições. Mais ao meio da figura, temos três agrupamentos próximos envolvendo a maior parte dos substantivos, grupos contendo 89 e 279 elementos, e outro contendo em grande parte adjetivos, verbos, advérbios e substantivos (109 elementos). Estes três grupos capturam uma propriedade distribucional importante no PB: a

11 De fato, em termos aproximados, o pronome "lhe" é bem menos frequente: o "te" ocorre cinco vezes mais, o "me" dez vezes mais e o "se" vinte vezes mais, em nossos dados de aquisição (incluindo fala adulta e da criança).

possibilidade de todos estes elementos ocuparem a função de núcleo nominal. Finalmente, os três agrupamentos da parte superior, contendo 45, 56 e 108 elementos, respectivamente, capturam em grande parte elementos que, apesar da natureza diversa (verbos imperativos, advérbios locativos, interjeições, pronomes etc.), podem ocorrer isoladamente em enunciados, no contexto apropriado (em perguntas, como "ali?", ou respostas, como "você!", por exemplo). O método distribucional evidencia assim uma propriedade importante da linguagem: há aspectos discursivos e propriedades pragmáticas que atravessam e se manifestam através de diversas classes gramaticais. Vê-se, também, porque a informação distribucional das palavras é, por si só, insuficiente para categorizá-las: a distribuição de palavras nos enunciados não expressa apenas propriedades sintáticas.

Figura 3 - Agrupamentos de elementos funcionais que são, a maioria, de natureza nominal e pronominal, contendo 45 e 32 itens, respectivamente.

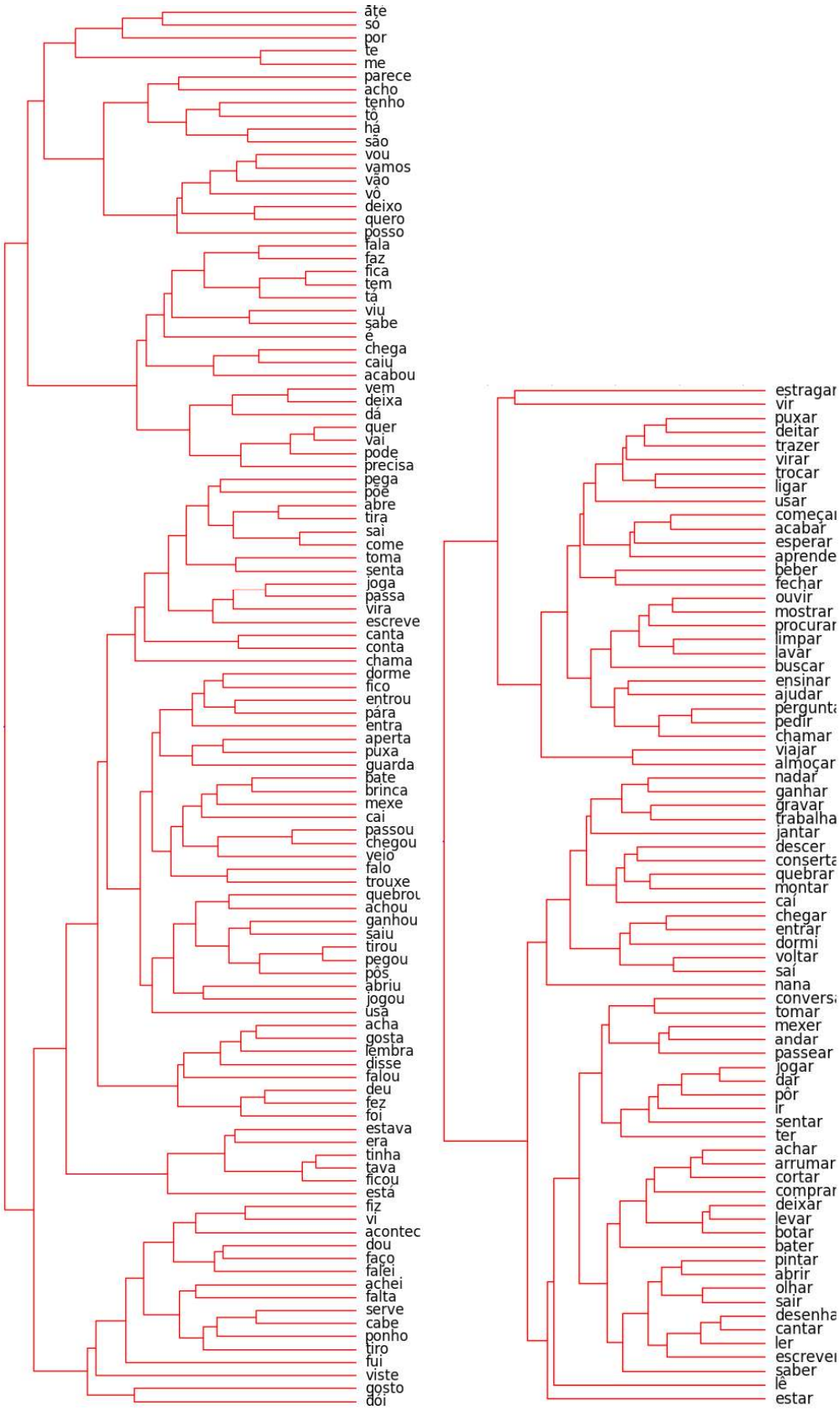


Nos voltamos agora para a análise de alguns agrupamentos específicos. A Figura 3 exibe dois agrupamentos de elementos de natureza funcional, em sua maioria. A distribuição à esquerda captura relativamente bem certas classes de elementos, tais como, por exemplo, os pronomes pessoais: vemos "eu" e "nós" formando um subgrupo, bem como "ele", "ela", "você", "vocês" e "eles" ou, ainda, "sua" e "minha". Vemos um subgrupo com "ainda" e "também" e outro com "já", "não" e "se". Note, no caso deste último, que os três elementos são bastante comuns em posição imediatamente pré-verbal. A distribuição captura ainda vários pares envolvendo apenas variação de gênero, como "do/da" e

"desse/dessa", ou variação dêitica, como "esses/aquele" e "este/isto". Note, ainda, que pronomes interrogativos ficaram muito próximos: "aonde, onde, quem, que, quando, como". No agrupamento à esquerda, observa-se padrões semelhantes: "algum, cada, quantas, nenhum" bem próximos, assim como "tuas, teus", "aqueles, estes, estas, aquelas" e "às, nas, à".

Estes dois agrupamentos apresentam poucos elementos "intrusos": os itens "filho", "será", "pois", "cheio" e "causa". Estes não parecem se adequar bem ao demais elementos. Mas de modo geral, os grupos consistem de pronomes pessoais, possessivos, demonstrativos, interrogativos e clíticos; advérbios, quantificadores, contrações de preposições com artigos e pronomes, e interjeições. Olhando para o dendrograma da Figura 2, porém, vemos que os dois grupos da Figura 3 estão bastante separados: o de 45 elementos está na parte de cima, mais próximo dos agrupamentos envolvendo substantivos , enquanto o de 32 está na parte de baixo, mais próximo dos agrupamentos verbais. Embora possa ser afetada significativamente pelo tamanho do corpus e também pela consistência de sua transcrição, uma possível explicação para essa distribuição é a de que o agrupamento da esquerda deve envolver elementos que, no corpus FDS, ocorrem circundados por elementos do domínio verbal (como advérbios, pronomes acusativos e oblíquos, etc.), enquanto o agrupamento da direita envolve elementos que aparecem circundados por itens do domínio nominal, daí a preponderância de pronomes possessivos e contrações de preposições com artigos.

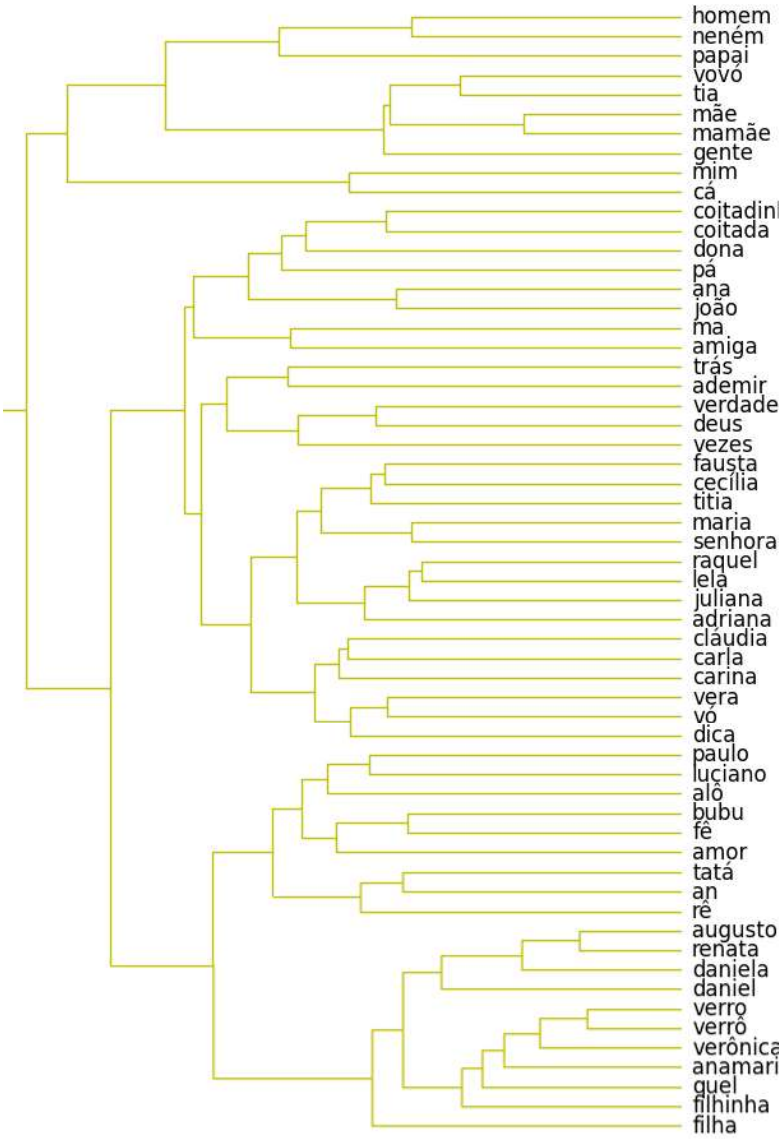
Figura 4 - Agrupamento "verbal" contendo 108 elementos, dos quais 95 são verbos flexionados (à esquerda). Um segundo agrupamento de 74 elementos, dos quais 73 são verbos no infinitivo.



Na Figura 4, vemos dois dos 4 agrupamentos caracterizados majoritariamente por elementos verbais. O agrupamento da esquerda contém 95 verbos flexionados presente ou passado. Vemos aí também alguns subgrupos que capturam importantes propriedades distribucionais dos verbos. Por exemplo, um subgrupo composto por "estava, era, tinha, (es)tava, ficou, está", todos verbos com muitas semelhanças em suas funções sintáticas em construções predicativas ou em locuções verbais. A forma "tava" aparece aí porque optamos, como em Redington *et al.* (1998), por não normalizar a ortografia do corpus. Em parte, isso se deve ao fato de que há variação na fala envolvendo alguns verbos, como é o caso de "estava" e "tava". Seria uma tarefa da criança, a priori, reconhecer essa variação e saber quando ela está diante dos mesmos verbos. A análise distribucional mostra que aí também reside uma fonte de informação, pois reflete isso na proximidade de itens como estes e "vô/vou". Ressaltamos também o subgrupo formado por "vou, vamos, vão, vô", que indica também a utilidade da informação distribucional no aprendizado das várias flexões de um mesmo verbo, no caso "ir".

O agrupamento da direita, por sua vez, inclui basicamente verbos no infinitivo, capturando muito bem sua distribuição, embora haja outras formas no infinitivo compondo outros agrupamentos. Nesse grupo, temos apenas uma exceção, que seria o substantivo "jantar" (formas como "lê" e "saí" que aparecem aí seriam registros da fala coloquial). Note, porém, que neste caso é um problema de homonímia entre o verbo "jantar" e o substantivo respectivo. Como no CTB o substantivo ocorre mais frequentemente, o método acaba "errando" com relação à categoria de referência, pois provavelmente o verbo é a mais frequente no corpus FDC. Esse caso mostra a importância de construir, no futuro, uma anotação de referência para o próprio corpus FDC, de modo que o método seja avaliado mais precisamente. De todo modo, o grau de "pureza" deste grupo ressalta em relação aos demais e indica que a experiência da criança é muito clara a seu respeito. Isso poderia ser uma das razões para que em línguas como o PB, crianças não apresentam claramente o estágio conhecida como de "infinito raiz", isto é, por orações simples com verbo no infinitivo, como ocorre com crianças adquirindo o inglês.

Figura 5 - Uma parte do agrupamento de substantivos contendo 279 elementos.

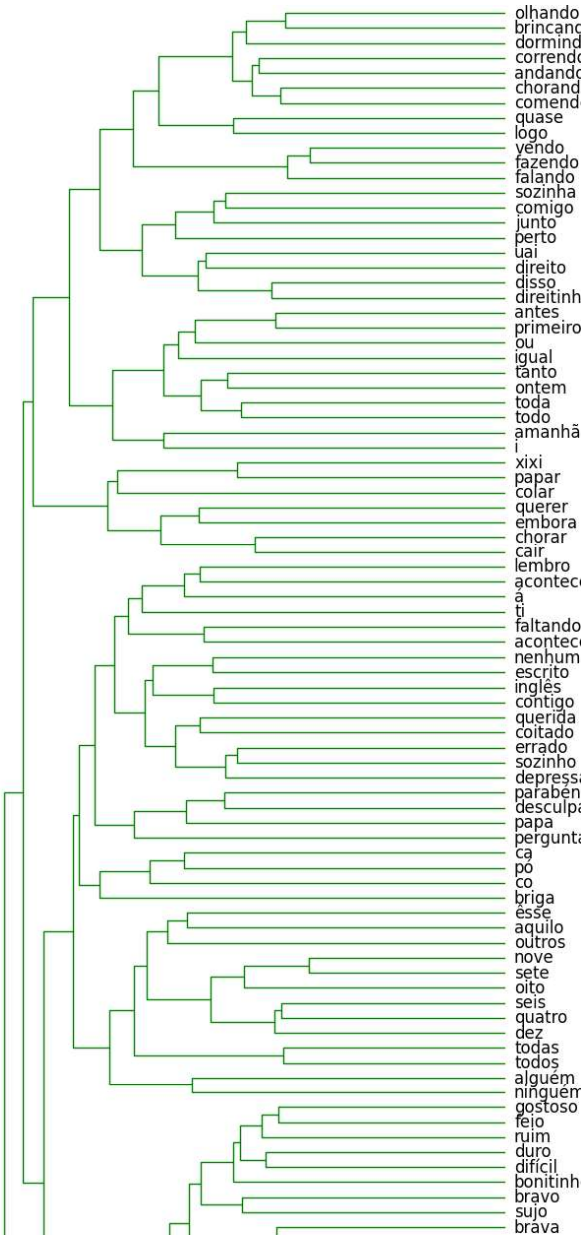


O maior agrupamento, como esperado, é o de 279 elementos parcialmente exibido na Figura 5, com 254 substantivos (ou seja, uma "pureza" de 91%). Neste grupo, se misturam nomes comuns e nomes próprios. Porém, é possível ver que há alguma tendência a aglomerar nomes próprios, como vemos no subgrupo exibido acima. Note que mesmo ocorrências normalmente tidas como nomes comuns, como "papai", "vovó" e "neném" possuem uma característica de nome próprio na FDC, pois em geral se referem a pessoas bastante específicas e podem ser tomados pelo próprio nome das mesmas. Novamente, como se vê aqui, não houve normalização dos nomes, razão pela qual se vê variantes como "verro" e "verrô" ou "mãe" e "mamãe". Assim como no caso dos verbos, esperava-se que os substantivos fossem capturados de modo mais preciso pelo método e isso se confirmou, o que tende a beneficiar a cobertura. Como dito anteriormente, foi por esta razão que optamos por

privilegiar a precisão geral no cômputo da medida *F*.

Finalmente, temos na Figura 6 uma amostra parcial de um agrupamento formado por adjetivos, verbos e advérbios. No caso dos verbos, vemos na parte de cima uma certa quantidade de verbos no gerúndio. É possível que seu aparecimento juntamente com os adjetivos se dê por uma distribuição parecida em construções com verbos auxiliares, como "ser" e "estar", por exemplo, que também tomam adjetivos e mesmo advérbios como complementos ou modificadores adjacentes (por exemplo, "ele estava fazendo/cansado/ontem ..."). O agrupamento conta ainda com um bom número de substantivos (15) e vemos também outras categorias nesse agrupamento, tais como numerais ("dez, seis, sete" e outros) e quantificadores ("todo, toda, todos todas") e ainda a conjunção "ou", além de outros elementos residuais. Como se percebe, há um grau relativamente alto de heterogeneidade nos agrupamentos envolvendo outras categorias que não a verbal e a nominal. De todo modo, os agrupamentos apresentados aqui demonstram que várias relações de similaridade categorial entre as palavras foram capturadas pelo método e, como vemos nos gráficos de desempenho, sempre em um grau bastante superior ao do acaso, o que corrobora a assunção de que a informação distribucional é útil para a criança na aprendizagem das categorias. No que segue, apresentamos os resultados quantitativos de cada experimento realizado neste estudo.

Figura 6 - Uma parte do agrupamento de 109 elementos, contendo principalmente adjetivos, verbos e advérbios.

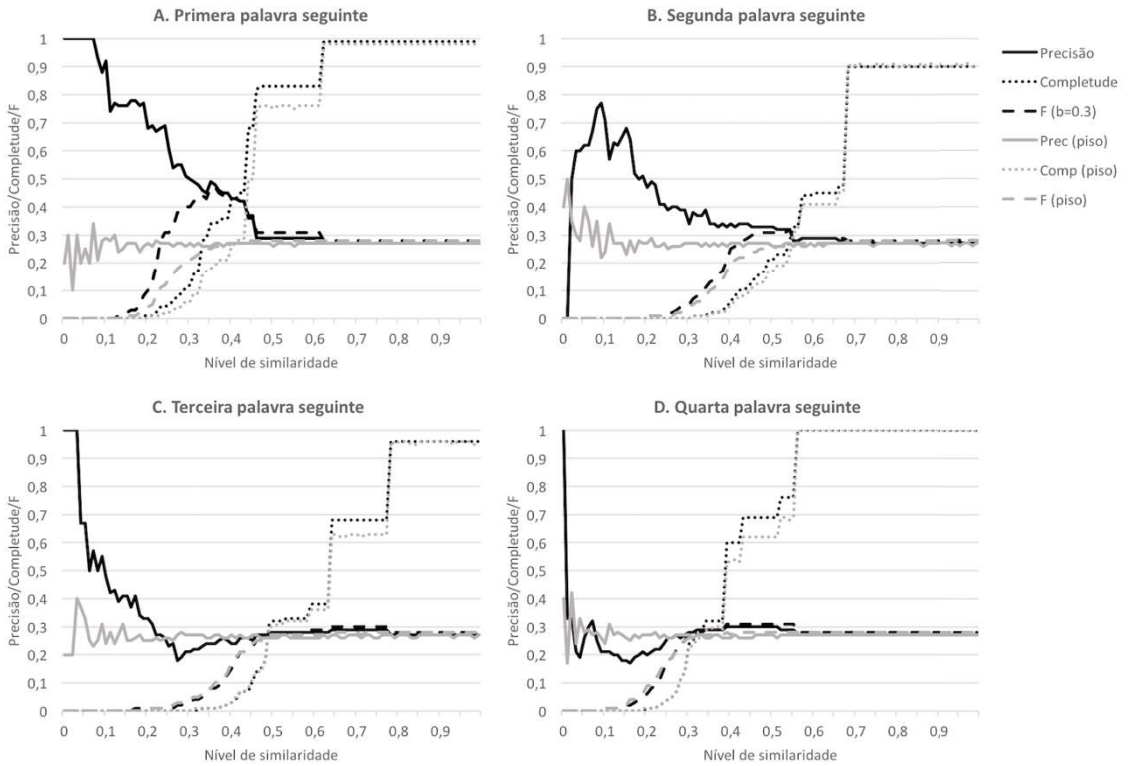


4.2. Experimentos adicionais: variando os contextos

Neste experimento, o objetivo era averiguar a informatividade dos contextos precedente e posterior à palavra-alvo, de modo a saber qual o mais informativo e se a distância do item de contexto em relação à palavra-alvo também afeta a informatividade. No experimento padrão, usou-se a janela de contexto que inclui as duas palavras imediatamente precedentes e também as duas imediatamente seguintes. A partir dos resultados apresentados a seguir, tem-se uma ideia mais precisa da utilidade da

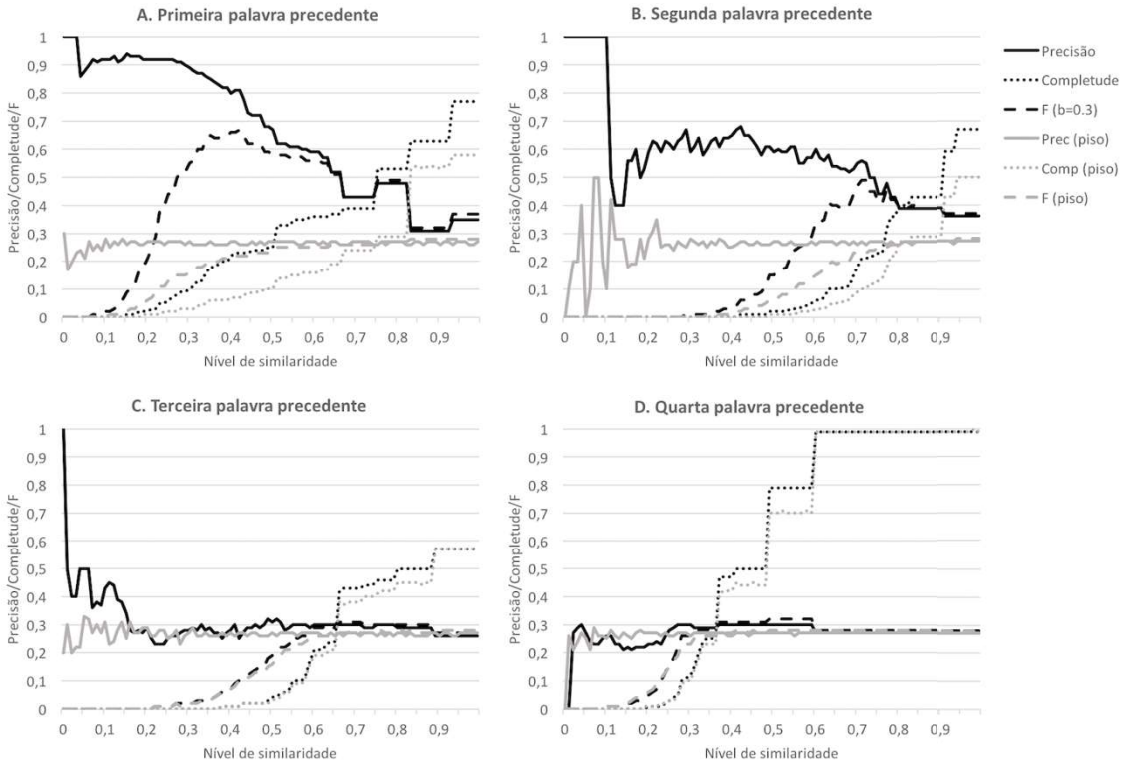
informação distribucional para a criança. As medidas obtidas estão compiladas na Tabela 2, ao final desta seção.

Figura 7 - Desempenho quando a primeira (A), segunda (B), terceira (C) e quarta (D) palavras seguintes\ **são usadas como contexto. As linhas negras representam o desempenho do método, enquanto as** linhas acinzentadas o piso classificatório.



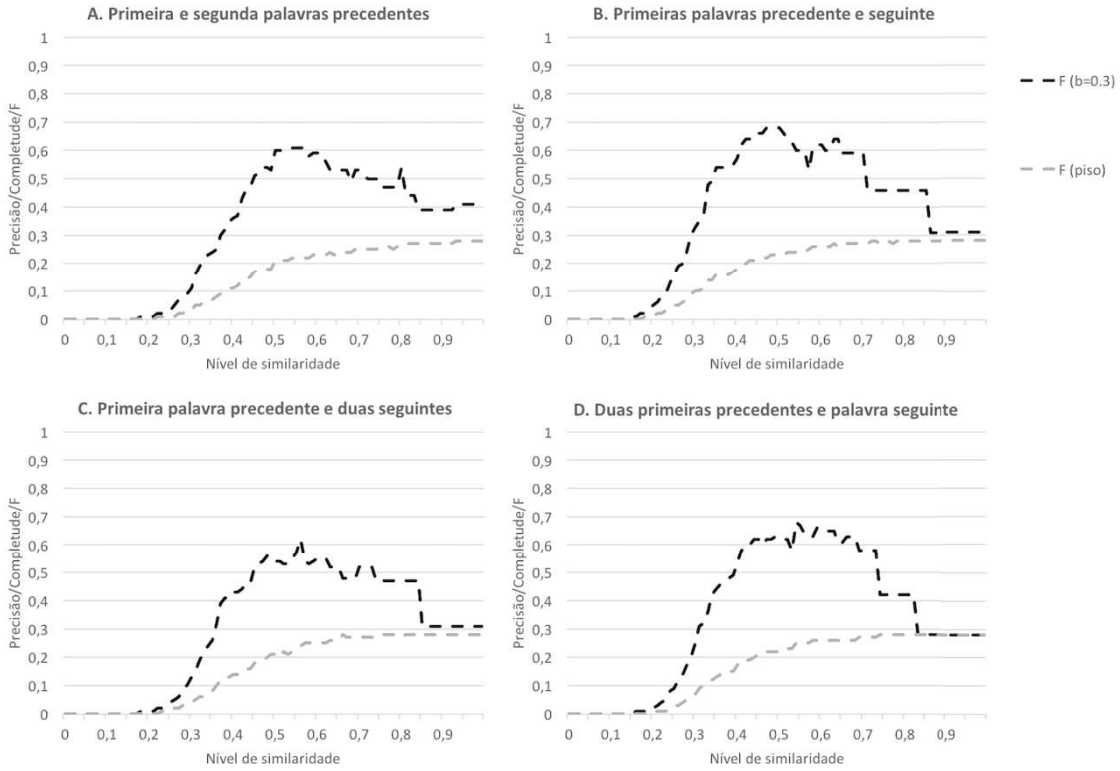
Começamos pela avaliação do contexto que sucede a palavra-alvo, como mostra a Figura 7 (A, B, C e D). São exibidas todas as medidas de desempenho (precisão, completude e *F*), tanto para o método, quanto para o piso classificatório, produzido a partir da média de 10 classificações geradas aleatoriamente, para cada condição. Com relação à posição do contexto, fica evidente que à medida que a distância aumenta, diminui sua informatividade para a classificação da palavra-alvo ao ponto de podermos concluir que apenas a primeira palavra (Figura 7A) seguinte é realmente informativa (*F*=0,48, corte em 0,37, 19 grupos), enquanto as demais pouco ou nada contribuem, quando comparadas ao piso classificatório. Estes resultados são muito similares aos do inglês, apresentados em Redington *et al*. (1998), o que indica que as duas línguas se comportam de modo similar nesse aspecto.

Figura 8 - Desempenho quando a primeira (A), segunda (B), terceira (C) e quarta (D) palavras precedentes são usadas como contexto. As linhas negras representam o desempenho do método, enquanto as linhas acinzentadas o piso classificatório.



Após avaliar o contexto posterior, foi avaliado também o contexto precedente. A Figura 8 (A, B, C e D) apresenta os resultados. Da mesma forma, vê-se claramente que quanto mais próximo, mais informativa é um item de contexto para a categorização da palavra-alvo. Porém, o contexto precedente se mostra mais informativo que o posterior: neste, tanto a primeira quanto a segunda palavra precedente são informativas, embora a primeira seja a que carrega mais informação (F=0,67, corte em 0,41, 47 grupos). A terceira e a quarta palavras precedentes não são informativas. Vale ressaltar que, apesar do valor relativamente alto de F, na condição 8A foi a precisão (0,81) que "puxou" F para cima, visto que a cobertura é bastante baixa (0,23). Isso fica evidente quando se considera o alto número de agrupamentos: 47 (contra as 10 categorias definidas na classificação de referência).

Figura 9 - Desempenho (apenas medidas F) para contextos contendo mais de uma posição. A linha acinzentada é o piso classificatório.



Finalmente, no último bloco de simulações foram avaliadas janelas de contexto compostas por mais de um elemento, como mostra a Figura 9 (A, B, C e D), em que apenas as medidas *F* são exibidas. Note que o desempenho é semelhante em todas as condições, muito em função da presença da primeira palavra antecedente em todas elas que, como observado acima sobre a condição 8A, é a posição com maior grau de informatividade. A segunda posição que parece agregar mais informação é a primeira posição seguinte à palavra-alvo: as condições 9B e 9D são as que obtém os melhores resultados. Contrastando as condições 9B, 9C e 9D vemos ainda que a segunda palavra antecedente contribui para um aumento da cobertura do método, que chega ao máximo em 9D, com *F*=0,40 e apenas 14 agrupamentos, quantitativamente próximo aos 10 previstos na classificação de referência. Inversamente, como mostra a condição 9C, a segunda palavra seguinte parece atrapalhar a classificação, que cai de *F*=0,68 em 9B para *F*=0,62 em 9C. Portanto, os resultados indicam que para o português brasileiro o contexto distribucional ideal parece ser o que inclui as duas palavras imediatamente precedentes e a palavra seguinte à palavra-alvo. Inclusive melhor que o contexto utilizado no experimento padrão.

Tabela 2 - Medidas específicas obtidas para os vários contextos avaliados. Os valores da coluna "Contexto" indicam as posições relativas à posição da palavra-alvo.

Figura	Contexto	F	Precisão	Completude	Corte	Grupos
7A	[1]	0,47	0,49	0,34	0,35	24
7B	[2]	0,32	0,32	0,28	0,54	12
7C	[3]	0,30	0,29	0,68	0,64	3
7D	[4]	0,31	0,30	0,60	0,39	5
8A	[-1]	0,67	0,81	0,23	0,42	43
8B	[-2]	0,49	0,56	0,21	0,71	26
8C	[-3]	0,31	0,30	0,43	0,67	9
8D	[-4]	0,32	0,30	0,79	0,49	4
9A	[-2,-1]	0,61	0,75	0,20	0,54	42
9B	[-1,1]	0,68	0,79	0,27	0,48	33
9C	[-1,1,2]	0,62	0,68	0,33	0,56	23
9D	[-2,-1,1]	0,68	0,72	0,40	0,54	14

Considerações finais

Neste trabalho, apresentamos resultados parciais de uma investigação computacional em andamento, que visa replicar, para o português brasileiro, o estudo de Redington *et al*. (1998), bem como outros estudos da literatura, além de novas questões que estamos formulando. Foram apresentados aqui os experimentos que visam estabelecer o grau de informatividade do contexto distribucional, para diversas configurações do mesmo. Foi possível estabelecer, por exemplo, que a palavra imediatamente precedente à palavra-alvo é a que mais informa sobre sua categoria. Porém, a melhor performance classificatória foi obtida quando tomamos as duas palavras imediatamente precedentes mais a primeira palavra seguinte à palavra-alvo.

Verificou-se, ademais, que o contexto local (até duas palavras antes ou depois da palavra-alvo) é

informativo em grau muito acima do piso classificatório que consiste de uma categorização aleatória. Tais resultados indicam ser plausível supor que a criança extrai informação do contexto de uma palavra, como forma de acessar sua classe e, assim, seu significado. O fato de ter bastante informação em uma janela tão curta indica que esse tipo de exploração do dado de entrada é compatível com a hipótese da alavancagem prosódica (Christophe *et al*., 2008), na medida em que seria compatível com a extensão das frases fonológicas.

Com relação ao estudo original para o inglês de Redington *et al.* (op.cit.), há ainda outras condições experimentais que temos investigado, tais como variar o número de palavras-alvo e de palavras de contexto, checar para quais classes a informação distribucional é mais valiosa, variar o tamanho do corpus de entrada, avaliar o efeito de diferentes assunções sobre fronteiras de enunciado, avaliar se um corpus de fala entre adultos é mais informativo do que a fala dirigida à criança, entre outros. Portanto, em breve esperamos publicar novos resultados, complementares aos que aqui disponibilizamos. Ademais, a cada dia nos ocorrem novas perguntas relacionadas ao problema da aprendizagem distribucional, ainda inéditas na literatura da área e que pretendemos investigar.

Finalmente, há muitas questões importantes a debater sobre este tipo de modelagem, que envolvem desde aspectos puramente técnicos até aspectos que envolvem a discussão necessária sobre a plausibilidade psicológica e empírica de tais modelos. Por exemplo, vale ressaltar que o aprendiz modelado assume uma experiência instantânea, nos moldes das idealizações feitas por Noam Chomsky, quando formaliza aspectos da aquisição da linguagem. Será necessário, portanto, que no futuro o modelo seja capaz de capturar a aquisição gradual, a partir da exposição progressiva aos dados de entrada. Só aí o modelo atenderá a mais critérios psicolinguísticos que o tornem uma fonte fidedigna de informações sobre o processo de aquisição da linguagem pela criança.

REFERÊNCIAS

BERNAL, S.; LIDZ, J.; MILLOTTE, S; CHRISTOPHE, A. Syntax constrains the acquisition of verb meaning. *Language Learning and Development,* 3, p.325–341, 2007.

BERWICK, R. C.; PIETROSKI, P.; YANKAMA, B.; CHOMSKY, N. Poverty of the stimulus revisited. *Cogn Sci*, 35(7), p.1207-42, Sep-Oct 2011. DOI: 10.1111/j.1551-6709.2011.01189.x.

BROWN, R. Linguistic determinism and the part of speech. J*ournal of Abnormal & Social Psychology,* 55, P.1-5, 1957.

CHOMSKY, N. *Knowledge of Language:* its Nature, Origin, and Use. New York: Praeger, 1986.

EVANS, N.; LEVINSON, S. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences,* 32(5), p.429-448, 2009. DOI:10.1017/S0140525X0999094X.

HARRIS, Z. S. Distributional structure. *Word*, 10(2-3), p.146-162, 1954.

HOHLE, B.; WEISSENBORN, E.; KIEFER, D.; SCHULZ, A.; SCHMITZ, M. Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy*, 5(3), p.341-353, 2004.

KAPLAN, F.; OUDEYER, P. Y.; BERGEN, B. Computational models in the debate over language learnability. *Infant and Child Development*, 17(1), p.55-80, 2008.

LANDAU, B.; GLEITMAN, L. R. *Language and experience:* evidence from the blind child. Cambridge, MA: Harvard University Press, 1985.

MACWHINNEY, B. *The CHILDES Project:* Computational Tools for Analyzing Talk; Version 0.8. European Science Foundation, 1989.

MINTZ, T. H.; NEWPORT, E. L.; BEVER, T. G. The distributional structure of grammatical categories in speech to young children. *Cognitive Science,* 26(4), p.393-424, 2002.

NAIGLES, L. R. Children Use Syntax to Learn Verb Meanings. *Journal of Child Language*, 17(2), p.357-374, 1990.

PEARL, L. Using computational modeling in language acquisition research. *Experimental methods in language acquisition research*, 27, p.163, 2010.

PETERS, A. M. Early syntax. *Language acquisition,* 2, p.307-325, 1986.

PINKER, S. Formal Models of Language Learning. *Cognition*, 7, p.217-283, (1979.

PINKER, S. *Language learnability and language learning.* Cambridge, MA: Harvard, 1984.

REDINGTON, M.; CHATER, N.; FINCH, S. Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive science,* 22(4), p.425-469, 1998.

ROSSI, N. F.; MORETTI-FERREIRA, D.; GIACHETI, C. M. Genética e linguagem na síndrome de Williams-Beuren: uma condição neuro-cognitiva peculiar. *Pró-Fono Revista de Atualização Científica,* Barueri (SP), v. 18, n. 3, p.331-338, set.-dez. 2006.

YANG, C. Computational models of syntactic acquisition. Wiley Interdisciplinary Reviews. *Cognitive Science*, 3(2), p.205-213, 2011.