

Função na língua, generalização e reprodutibilidade

Raquel Freitag

Julian Tejada

Bruno Pinheiro

Paloma Cardoso

Universidade Federal de Sergipe (UFS)

Universidade Federal de Sergipe (UFS)

Universidade Federal de Sergipe (UFS)

Universidade Federal de Sergipe (UFS)

Neste texto, uma técnica para validar a intuição e sensibilidade do analista que codifica valores de uma função é apresentada. A técnica de aprendizagem de máquina por algoritmo de classificação é aplicada a dois conjuntos de dados de fenômenos polissêmicos do português brasileiro – funções dos diminutivos e do modalizador parentético epistêmico (*eu acho que* – para testar se os critérios adotados por um analista são consistentes e podem ser generalizados. Os resultados apontam para a importância da análise exploratória e confirmatória na rotina de classificação, e o aprendizado de máquina permite a identificar quando as condições de um modelo não são as responsáveis pela atribuição do julgamento.

Introdução1

O termo *função* em linguística é essencialmente polissêmico. Nichols (1984) descreve cinco diferentes acepções para este termo, somente no campo da terminologia linguística. Uma destas acepções é a de *função* = *significado*, que é frequentemente adotada em estudos de descrição linguística de fenômenos em níveis gramaticais mais altos, como o semântico-discursivo e/ou semântico-pragmático. Nestes estudos, uma dada forma (*item linguístico, construção*, ou outro rótulo, a depender do viés teórico de base funcionalista assumido) é associada a uma ou mais funções, configurando situações de polissemia ou de multifuncionalidade, ou uma mesma função/significado é passível de ser associada a mais de uma forma, configurando diversidade formal em um domínio funcional (TORRES-CACOULLOS, 2001).² Em uma abordagem integrativa, fatores de natureza funcional – decorrentes do pareamento entre forma e função – e fatores contextuais³ são responsáveis pela seleção de uma dada estrutura linguística na situação de uso, o que desencadeia a variação (CROFT, 2002).

Para ilustrar, tomem-se como exemplo elementos da língua denominados de *marcadores discursivos*, que são caracterizados por sua forma e por sua função. Enquanto os aspectos formais são sistemáticos e associados a uma posição específica no turno oracional e padrões prosódicos característicos, a estimativa da função é bastante diversificada. Por exemplo, os marcadores discursivos *tá?* e *certo?* (FREITAG, 2001) têm como funções: i) expressão de teste do canal com o interlocutor, ii) solicitação da sua aquiescência, iii) ordenamento e reorganização do fluxo conversacional, e iv) todas as funções anteriores ao mesmo tempo. Além disso, outros marcadores discursivos também podem desempenhar a mesma função múltipla de *tá?* e *certo?*, como é o caso do marcador discursivo *né?* (FREITAG; SILVA; EVANGELISTA, 2017), uma possibilidade que é condicionada pelo contexto. Considerando a diversidade formal, multifuncionalidade, ou polissemia de elementos linguísticos nestes níveis gramaticais, a sistematização de usos recai sob o juízo de valor de um analista, que atribui rótulos para cada função/significado, que são associados a uma dada forma, baseado primeiramente em suas intuições de falante da língua (conhecimento linguístico implícito), para depois testar a associação da função/significado atribuída a um conjunto de pistas contextuais (seja de natureza estrutural, subjetiva, cognitiva ou social), a fim de estabelecer os parâmetros de recorrência (conhecimento explícito). É, em linhas gerais, seguindo este percurso metodológico que os estudos descritivos têm sido desenvolvidos, dos quais são derivados conhecimentos sobre a gramática do português brasileiro em níveis mais altos, como as

funções de marcadores discursivos (FREITAG, 2001; FREITAG; SILVA; EVANGELISTA, 2017), as funções da forma de pretérito imperfeito (FREITAG, 2011), funções de *tipo* (SANTANA, 2019), dentre outras.⁴

Os resultados da sistematização proposta para os significados de uma função/forma por um analista em uma dada amostra temporal, espacial e socialmente situada, são tomados como base para outros estudos, em busca de uma generalização do comportamento. Voltando ao estudo dos marcadores discursivos, diferenças no comportamento de marcadores discursivos interacionais *entendeu?, sabe?, né? e certo?* foram encontradas comparando resultados de estudos em amostras que não compartilhavam a dimensão temporal e espaço-dialetal (FREITAG; SILVA; EVANGELISTA, 2017), os resultados podem ser consequência dessa diferença, mas também podem ser decorrentes de decisões individuais dos próprios analistas. Este cenário compromete a potencialidade de generalização e até mesmo a reprodutibilidade da ciência, uma vez que a definição das funções é uma tarefa de decisão do analista, que, por vezes, encontra dificuldades para definir o sentido/significado, sugerindo funções ambíguas, como a ilustrada em iv) acima. Assim, em vez de caracterizar as funções que uma dada forma pode codificar na língua por meio da associação a pistas de contextualização, o objetivo de estudos como os ilustrados anteriormente poderia ser mais adequadamente formulado como a aferição dos efeitos das pistas de contextualização na decisão do analista em caracterizar funções de uma dada forma. Com esta formulação, o poder explanatório é reduzido, limitando a generalização dos resultados.

No entanto, este *modus operandi* é muito anterior à pesquisa descritiva linguística, remontando à base da gramatização das línguas: é o gramático, baseado em sua intuição e sensibilidade, que codifica a regra. Do mesmo modo, tem sido o linguista-analista, baseado na sua intuição e sensibilidade, que examina os usos e que atribui uma função/significado à forma. Este *modus operandi* parece ser uma característica das humanidades de modo geral, o que leva muitas vezes à sensação de que a pesquisa neste campo é baseada em intuições de pesquisador/analista, imbuído em sua experiência individual, o que dificultaria *a priori* a replicação de resultados de uma investigação. Neste texto, apresentamos uma proposta que visa contribuir não só para a replicabilidade (entendida aqui como a replicação de um estudo com o mesmo conjunto de dados e a mesma abordagem analítica) como também para a generalização (entendida aqui como a mesma abordagem analítica em conjuntos de dados diferentes) das sistematizações de descrição linguística. Em conjuntos de dados de dois fenômenos polissêmicos do português brasileiro – funções dos diminutivos e do modalizador parentético epistêmico (*eu acho que* – aplicamos a técnica de aprendizagem de máquina para testar se os critérios adotados por um analista são consistentes e podem ser assumidos para estudos de generalização.

1. Fenômenos

1.1 Diminutivos

O processo de derivação de nomes primitivos com os sufixos [x-inho] e [x-zinho] para a função de indicar tamanho pequeno é conhecido na tradição gramatical do português como diminutivo: *casa – casinha, pato – patinho, gato – gatinho*. No entanto, outras funções/significados começam a ser descritas nas gramáticas associadas aos diminutivos, como afeição, compaixão, intensidade, quantidade, associadas ao domínio da afetividade. A expansão de funções/significados codificada nas gramáticas foi gradual; no decorrer dos séculos XVIII-XX, os compêndios gramaticais ampliaram os sentidos atribuídos ao sufixo de diminutivo. Em levantamento nestes instrumentos, Santana (2017, p. 37) identificou:

1. Tamanho pequeno: *casinha* (casa pequena) e *carrinho* (carro pequeno);
2. Aproximação afetiva positiva: *caminha* (cama agradável) e *corpinho* (corpo bonito);

3. Depreciação: *fradezinho* (frade de pouca importância) e *gentinha* (grupo de pessoas sem valores morais);
4. Intensidade: *friozinho* (muito frio), *devargazinho* (muito devagar);
5. Duração: *chegadinha* (chegar a um determinado lugar) e *instantinho* (instante);
6. Quantidade: *aguinha* (pouca água) e *bigodinho* (pouco bigode).

A morfologia de diminutivo também gera novos itens na língua, resultado de processo de lexicalização, em que o valor da base é modificado, a exemplo de *folhinha* (referência ao calendário, e não uma folha pequena) e *coxinha* (referência ao salgado, e não uma coxa pequena), considerados “falsos diminutivos” por não estarem associados a um sentido de tamanho pequeno ou um sentido afetivo da base, mas a um novo significado referencial (ROCHA; VICENTE, 2016).

A forma de diminutivo ([x-inho] ~ [x-zinho])⁵ é polissêmica, significando tamanho pequeno, expressando valores afetivos, ou se encaminhando para a formação de outro item lexical independente da base. E essa polissemia é registrada nas gramáticas, que passaram a registrar não somente diminutivos que indicavam tamanho pequeno, mas também começaram a descrever valores afetivos, de apreciação tanto de forma positiva como de forma negativa, associados à forma de diminutivo. Muitas vezes, os usos de dimensão pequena (uso reduzido) carregam marcas afetivas por envolverem sentimento ou memória afetiva. Os diminutivos afetivos carregam marcas de apreciação em relação ao seu referente, essas marcas podem ser positivas (julgamentos positivos) ou negativas (julgamentos negativos).

A identificação da função do diminutivo é feita pelo analista, caso a caso, que julga o tipo de função associada à morfologia do diminutivo no contexto linguístico sob análise. Considerando que a morfologia do diminutivo no português pode ter função lexicalizada ou não, e que as funções não lexicalizadas podem ter valores afetivos de apreciação positiva ou negativa, e que estes julgamentos de valor são realizados pelo falante/analista no contexto linguístico, é possível relacioná-los a pistas contextuais, de natureza estrutural e subjetiva, controladas em um modelo como variáveis independentes.

1.2. Modalizador (eu) acho que

Construções de verbos de cognição que encabeçam oração, como *I think* (THOMPSON; MULAC, 1991), no inglês, e *(eu) acho que* (CASSEB-GALVÃO, 2000; FREITAG, 2007) no português, cristalizadas na primeira pessoa do singular do presente do indicativo, são reanalisadas como modalizadores parentéticos epistêmicos, e indicam “atitude do falante e seu julgamento acerca da informação proposicional da oração” (FREITAG, 2007, p. 84), no domínio da modalidade epistêmica, expressando sentidos de certeza, incerteza e dúvida.

(Eu) acho que vai chover hoje.

‘tenho certeza’

‘é possível’

‘não tenho certeza’

A polissemia dos modalizadores parentéticos epistêmicos como *(eu) acho que* é resultado de mudanças semânticas do verbo *achar* em processo de gramaticalização, que impulsionam a expansão conceitual sofrida por um item cujo sentido vai do mais concreto ao mais abstrato (TRAUGOTT; DASHER, 2004). Segundo Furtado da Cunha, Oliveira e Votre (1999), a mudança semântica de *achar* segue a escala proposta por Heine, Claudi e Hünemeyer (1991): *corpo* ->

objeto -> processo -> espaço -> tempo -> qualificação. Em sua forma latina, *achar* (*afflare*), que significava soprar, fazia parte do repertório linguístico de caçadores. Ao longo das mudanças que sofreu, de soprar passou a significar sentir a proximidade da caça pelo odor, arejar e, posteriormente, descobrir ou encontrar a caça. Nesse percurso, de verbo intransitivo, achar passou a transitivo, selecionando um objeto direto. Seguindo o percurso, deixou de indicar algo concreto para indicar algo no mundo das ideias, abstrato. E, mudando de categoria, passou a ter funcionamento independente da oração, como parentético epistêmico.

Os diferentes tipos de *achar* são categorizados pelo analista (e inferidos pelos falantes da língua), a partir de inferências contextuais, o que resulta em diferentes funções para a mesma forma. Com base no controle de sete propriedades (paráfrase, variabilidade temporal, variabilidade de modo, pessoas do verbo, presença e tipo do argumento interno, tipo de sentença em que aparece, e caracterização), Casseb-Galvão (2000) propõe cinco funções para o verbo *achar*, considerando um contínuo que abrange *achar* como verbo pleno, *(eu) acho que* como elemento que introduz complementos oracionais, até *(eu) acho que* cujo comportamento sintático se assemelha a um advérbio (ou um parentético epistêmico). Freitag (2007) relaciona os usos de *(eu) acho (que)* a variáveis estruturais, como o aspecto da situação sob escopo (acabado, inacabado) e presença/ausência de *que* como complementizador, e subjetivas, como o envolvimento do falante (direto, mediano, indireto), sequência discursiva (narrativa de experiência de vida, opinião, explanação, descrição) e grau de complexidade do assunto (mais ou menos complexo, a depender da familiaridade do falante).

Do ponto de vista estrutural, no estudo de Casseb-Galvão (2000) o uso de *(eu) acho que* para indicar considerações e suposições apresentou um padrão específico: *acho + que*, com ocorrência de *eu* em contextos em que é identificada maior responsabilidade do falante quanto ao que foi dito; resultados semelhantes aos descritos no estudo de Freitag (2007). Apesar das diferentes finalidades - enquanto Casseb-Galvão (2000) descreveu um contínuo de gramaticalização para *achar*, Freitag (2007) descreveu os fatores influenciaram a variação entre *acho (que)* e *parece (que)* para expressar opinião e percepção - ambos os estudos corroboram o comportamento de *(eu) acho que* como modalizador parentético epistêmico e ambíguo quanto aos valores de modalidade, expressando certeza, dúvida e incerteza. Dúvida e incerteza não são sinônimos: a primeira é uma alternância entre sim e não, que emerge a partir de experiências diretas, enquanto a segunda permite uma gradação justamente por estar relacionado às experiências indiretas (OLIVEIRA, 2011).

Em ambos os estudos (CASSEB-GALVÃO, 2000; FREITAG, 2007), a atribuição de sentidos ao modalizador *(eu) acho que* pelo analista foi associada a variáveis estruturais e subjetivas, a fim de identificar padrões de sistematicidade. No entanto, diferentes analistas propõem diferentes rótulos para as funções da mesma forma, o que torna a replicabilidade dos estudos sensível à subjetividade de cada analista, assim como no caso dos valores de apreciação dos diminutivos que são apresentados em gramáticas normativas.

2. Reprodutibilidade nas descrições linguísticas

A crise da ciência quanto à reprodutibilidade e replicabilidade da pesquisa não é uma novidade, mas a ênfase na viabilidade da reprodutibilidade nas humanidades (PEELS; BOUTER, 2018) parece ser ainda maior, como, por exemplo, evidenciam os desafios socioculturais na replicação em pesquisas de psicologia, e não é diferente na linguística.

A reprodução de uma tarefa é frequentemente difícil porque muitos aspectos que demandam adaptação podem influenciar o resultado. Mas, ainda mais frequentemente, o detalhamento completo de todo o procedimento, desde a descrição de ambiente até as intercorrências e particularidades de participantes, não é documentado ou descrito nos relatórios ou seções

metodológicas de produtos acadêmicos. Mesmo em campos da linguística em que a reprodutibilidade seria mais favorecida, como na linguística computacional, a disponibilização de códigos de análise é menos frequente do que a disponibilização do conjunto de dados. E, mesmo quando conjunto de dados e código são disponibilizados, não há garantias de que os resultados sejam reprodutíveis, como apontam Wieling, Rawee e Noord (2018).

Berez-Kroeker *et al.* (2018) defendem que práticas de reprodutibilidade na linguística são favorecidas no campo da documentação e descrição linguística, com o compartilhamento de dados linguísticos, na forma de bancos de dados digitais, o que permitiria que análises independentes fossem realizadas. No entanto, como mostramos no caso dos diminutivos e do modalizador parentético epistêmico (*eu acho que*), a reprodutibilidade de resultados depende da intuição e sensibilidade do analista, já que o procedimento metodológico envolve a construção de um modelo para identificar associação do valor da função atribuída pelo analista a pistas estruturais ou subjetivas, controladas como variáveis independentes.

Para ampliar o poder explanatório de descrições linguísticas, contribuindo com resultados mais robustos e realistas, Nicenboim *et al.* (2018) sugerem mudanças nas práticas metodológicas; uma delas é a separação entre análise exploratória e análise confirmatória. Os autores realizam este procedimento para um estudo sobre o processamento de sentenças no alemão, dividindo-o em duas etapas, uma exploratória e outra confirmatória, o que permite checar se os resultados obtidos na etapa exploratória são consistentes na etapa confirmatória.

A separação das etapas em um estudo descritivo, tal como nas funções dos diminutivos ou do modalizador parentético epistêmico (*eu acho que*), pode ser implementada com a divisão do conjunto de dados em um conjunto de dados de treino, para a análise exploratória, e um de teste, para a análise confirmatória. Esta é uma rotina em aprendizado de máquina (*machine learning*) em que a divisão costuma ser de 70%-80% e 30%-20%, aleatoriamente, para cada um dos respectivos conjuntos de dados (GOODFELLOW; BENGIO; COURVILLE, 2016).

Das rotinas de aprendizado de máquina, a aplicação de algoritmo de árvore de decisão aos conjuntos de dados, separados em um conjunto de treino e um conjunto de teste, pode auxiliar na identificação de padrões associados à classificação atribuída pelo analista para a função de cada um dos itens linguísticos sob análise. A aplicação de algoritmos de árvore de decisão para estudos exploratórios linguísticos tem se mostrado produtiva em abordagens sociolinguísticas, como o de *random forests* (SIMAKI; MPORAS; MEGALOOIKONOMOU, 2016; WALKER, 2020), ou inferências condicionais (FREITAG; PINHEIRO, 2020). Neste tipo de análise, um conjunto de dados classificados quanto a diferentes características (ou atributos) é submetido a um algoritmo de classificação para identificação de padrões de agrupamento: o algoritmo de classificação deduz estruturas a partir do modelo, subdividindo o conjunto de dados. Como em uma árvore, folhas se arranjam em galhos (nós), recursivamente, selecionando quais os atributos que melhor particionam o conjunto de dados. As árvores de decisão geradas pelo algoritmo C4.5 (QUINLAN, 1993) podem ser utilizadas para classificação, funcionando como classificadores estatísticos; J48 é uma implementação de código aberto em Java do algoritmo C4.5 no aplicativo de mineração de dados *Weka* (GARNER, 1995), que realiza testes para cálculo de ganho e entropia, que permitem a avaliação do modelo.

A atribuição pelo analista dos sentidos aos diminutivos e ao modalizador em um conjunto de dados autênticos configura-se como um procedimento de aprendizado supervisionado, em que um humano (analista) realiza a classificação do conjunto de dados, que é submetido ao algoritmo de classificação posteriormente.



Figure 1. FIGURA 1 - Atribuição de sentidos pelo analista e o fluxo de aprendizado supervisionado.

A decisão de considerar um diminutivo como lexicalizado ou afetivo (Objetivo₁), ou se é dúvida, incerteza ou certeza que (*eu*) *acho que* expressa, é uma ação do analista associada a um conjunto de pistas, controladas como variáveis independentes, ou, na terminologia da classificação, atributos ou condições. Nos modelos dos fenômenos sob análise, os atributos ou condições podem ser de natureza estrutural ou subjetiva. São de natureza estrutural, por exemplo, a forma de realização dos diminutivos (Condição₃) a classe gramatical (Condição₇), a realização da base do diminutivo (Condição₄), posição da sílaba tônica (Condição₅) e o número de sílabas da base (Condição₆), e a posição no enunciado e presença vs. ausência de complementizador para (*eu*) *acho que*. Já atributos ou condições de natureza subjetiva são relacionados à experiência do falante e inferidos pela intuição do analista a partir da imersão no contexto de produção, como o tipo de sequência textual (Condição₈), o envolvimento do falante com o assunto (Condição₉) ou o tópico discorrido (Condição₁₀). Além destas, existem variáveis que apresentam efeito aleatório (não previsível), como o participante (Condição₁) e o item lexical (Condição₂). Todas estas variáveis são categóricas; variáveis prosódicas, como a frequência fundamental, intensidade e duração, aferidas por meio de softwares específicos, são medidas contínuas, e que, por serem aferidas fisicamente, garantem maior possibilidade de replicação, desde que os parâmetros de medida sejam devidamente especificados.

Condição o1	Condição o2	Condição o3	Condição o4	Condição o5	Condição o6	Condição o7	Condição o8	Condição o9	Condição o10	Objetivo 1	Objetivo 1
fra3ff	barzinho	zinho	consoante	oxítona	uma sílaba	nome	narrativo	lazer	menor envolvimento	lexicalizado	
fra3ff	showzinho	zinho	consoante	oxítona	uma sílaba	nome	narrativo	lazer	maior envolvimento	afetivo	positiva
fra3ff	pertinho	inho	vogal média baixa	paroxítona	duas sílabas	advérbio	opinativo	educação	menor envolvimento	afetivo	negativa

car1ff	joguinho	inho	vogal média baixa	paroxítona	duas sílabas	nome	narrativo	moradia	maior envolvimento	afetivo	positiva
car1ff	negocinho	inho	vogal média baixa	paroxítona	três sílabas	nome	narrativo	moradia	maior envolvimento	afetivo	positiva

Table 1. QUADRO 1 - Condições e objetivos de uma classificação.

O objetivo da classificação é identificar quais atributos ou condições estão associados à decisão do analista para cada função, desvelando as regras subjacentes ao processo. Uma matriz de confusão alinha as frequências observadas e previstas e permite calcular parâmetros de avaliação do desempenho para cada classe, como a taxa de precisão, que, a partir da quantificação de classificações corretas e incorretas, afere o quanto o algoritmo acertou na classificação. Outro parâmetro utilizado para a avaliação de algoritmos de classificação que costuma ser adotado é o coeficiente de Kappa (κ), que afere a razão de chances na concordância de classificação correta entre amostras. Aplicação deste coeficiente para análises sociolinguísticas podem ser encontradas em Freitag (2019).

Japkowicz e Shah (2011) sugerem que a combinação de parâmetros de medidas de desempenho pode resultar em melhores avaliações dos algoritmos e, por tabela, dos modelos previstos pelos pesquisadores. Nessa perspectiva, adotamos a métrica SAR (CARUANA *et al.*, 2004), que combina erro quadrático (RMSE), acurácia (A) e a área abaixo da curva ROC (AUC), com a seguinte fórmula:

$$SAR = \frac{[A + AUC + (1 - RMSE)]}{3}$$

A comparação entre a taxa de precisão, o coeficiente e o parâmetro SAR na classificação do conjunto total e nos subconjuntos de treino e de teste para a validação do modelo pode resultar em um teste de confiabilidade nas análises de funções propostas por analistas, baseados em pistas dos atributos ou condições estruturais, subjetivas e prosódicas.

A natureza dos atributos pode interferir no desempenho da classificação: enquanto atributos ou condições estruturais têm maior potencialidade de serem aprendidos pelo algoritmo, gerando maior ganho, os subjetivos dependem de outras informações que não são previstas pelo modelo, gerando maior entropia.

No contexto deste estudo, propomos a classificação dos conjuntos de dados dos valores do modalizador (*eu acho que*) e dos valores dos diminutivos para avaliar se os critérios adotados por um analista são consistentes e podem ser assumidos para estudos de generalização, considerando uma hierarquização, do mais ao menos generalizável, medida pelos parâmetros de desempenho do algoritmo: variáveis estruturais e prosódicas teriam comportamento mais [+ generalizável], manifestado por maior taxa de precisão, grau de concordância (κ) ótimo e parâmetro alto, enquanto variáveis subjetivas teriam comportamento [- generalizável], decorrente da sensibilidade do analista, variando especialmente quanto ao grau de concordância.

Uma crítica aos algoritmos de classificação é a sensibilidade ao conjunto de dados, o que restringe o poder de generalização. Neste estudo, objetivamos investigar se a identificação dos atributos ou condições pelo algoritmo é influenciada pelo tamanho da amostra (com uma amostra exploratória de treino e uma amostra confirmatória de teste, como sugerem Nicenboim *et al.* (2018)) e pelo tipo de atributo (pistas contextuais controladas como variáveis estruturais, prosódicas ou subjetivas).

A hierarquização de critérios visa comparar o desempenho dos algoritmos classificadores resultantes de aprendizagem de máquina (em um processo de reamostragem) com o juízo de classificação do analista. Estudos que consideram a comparação de desempenho entre algoritmos classificadores resultantes de aprendizagem de máquina e humanos mostram que decisões de natureza subjetiva têm peso no resultado. Kononenko e Bratko (1991), ao comparar resultados de

algoritmos e de médicos em diagnósticos de câncer, destacam que os médicos, ao diagnosticar, não apenas tentam maximizar a precisão de sua resposta como também consideram outros fatores, como a influência de um diagnóstico na escolha do tratamento médico e o desconforto que causaria ao paciente. Tal cenário interfere na acurácia e na precisão do modelo, mas são fatores que não podem ser desconsiderados se o objetivo do aprendizado de máquina é simular o comportamento humano. Do mesmo modo, imaginamos ocorrer com a classificação de funções de fenômenos linguísticos.

3 Método

3.1 Conjunto de dados

A análise da classificação das funções dos diminutivos e de *(eu) acho que* toma como base o conjunto de dados dos estudos de Pinheiro (2021) e Cardoso (2021), respectivamente, disponibilizados no repositório *Open Science Framework* (, para os diminutivos, e , para os modalizadores). As ocorrências dos fenômenos foram extraídas de 30 entrevistas sociolinguísticas do *Banco de dados Falantes Sergipanos*, da Universidade Federal de Sergipe.

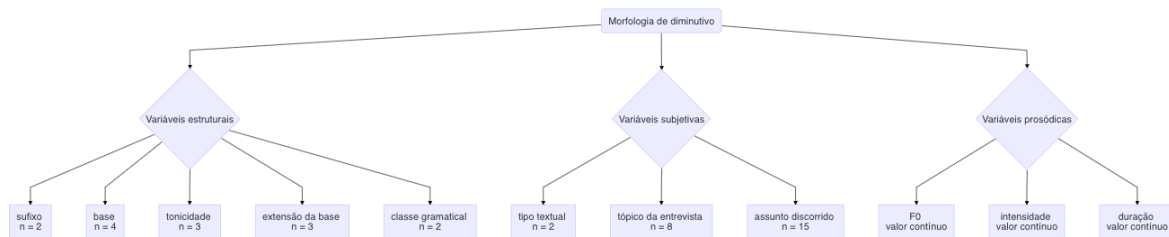


Figure 2. FIGURA 2 - Organização das variáveis do conjunto de dados de diminutivos.

Variáveis estruturais são as que podem ser inferidas a partir do contexto imediato do segmento da entrevista, por meio de categoriais previamente estabelecidas e convencionalizadas. As variáveis incluídas nos modelos dos fenômenos em questão foram selecionadas a partir de resultados de estudos anteriores.

Para os diminutivos, as variáveis estruturais controladas foram o sufixo ([x-inho] ~ [x-zinho]), o segmento final da base (vogal média-baixa [ɛ, ɔ], como em pé - pezinho, só - sozinho ~ vogal média-alta [e, o] como em cidade - cidadezinha, direito - direitinho ~ consoantes, como em bar - barzinho, final - finalzinho), tonicidade da base (oxítone ~ paroxítone ~ proparoxítone), extensão silábica da base (uma sílaba ~ duas sílabas ~ três ou mais sílabas), classe gramatical (nomes ~ outras classes).

Para os valores do modalizador *(eu) acho que*, as variáveis estruturais controladas foram o preenchimento da primeira posição (eu ~ 0), escopo do modalizador (sintagma ~ oração ~ turno), outro modalizador (presente ~ ausente), polaridade da oração (afirmativa ~ negativa).

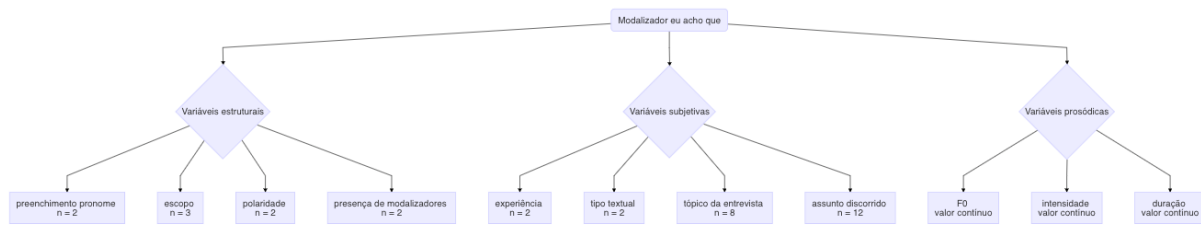


Figure 3. FIGURA 3 – Organização das variáveis do conjunto de dados do modalizador (eu) acho que.

As variáveis subjetivas *tipo de assunto*, *tópico discursivo* e *tipo de texto* foram controladas pareadamente para os dois fenômenos sob análise, e decorrem da inferência do analista a partir de pistas contextuais da oitiva das entrevistas sociolinguísticas. Existem diferentes propostas de pistas a serem consideradas na inferência do tipo do assunto, tópico e tipo de texto (FREITAG, 2014). No escopo da amostra considerada, o roteiro das entrevistas sociolinguísticas considerava oito tópicos de temas, que permitiam a emergência de diversos assuntos, em sequências de texto narrativas e opinativas. Especificamente para os valores do modalizador (*eu*) *acho que*, também foi controlada a experiência do falante (experiência direta ~ experiência indireta), inferida a partir da oitiva de segmento maior de entrevista.

O controle de variáveis prosódicas segue evidências de estudos que sugerem associação entre padrões prosódicos e valor de diminutivos afetivos (VALDERRAMA; PÉREZ; FUENTES, 2017), e a diferença entre certeza, incerteza e dúvida (ANTUNES; AUBERGÉ, 2015). Os áudios das entrevistas foram recortados no *Praat* (BOERSMA; VAN HEUVEN, 2001) e os parâmetros de duração, intensidade e F0 dos segmentos-alvo foram extraídos e normalizados em semitons.

3.2 Procedimento de classificação

O algoritmo J48 apresenta desempenho satisfatório em rotinas de classificação (CARVALHO; GUEDES, 2019; HASAN et al., 2016; MAIRESSE et al., 2007; STIGLIC et al., 2012), e fornece os valores para o cálculo do parâmetro SAR (JAPKOWICZ; SHAH, 2011). Originalmente implementado pelo aplicativo *Weka* (GARNER, 1995), uma coleção de algoritmos para mineração de dados, neste estudo utilizamos o pacote *Rweka* (HORNIK; BUCHTA; ZEILEIS, 2009) da plataforma R para implementar o algoritmo J48, por ser um procedimento que torna mais fácil a manipulação dos conjuntos de dados originais.

A primeira rotina de classificação considerou o efeito do tamanho da amostra no resultado da classificação. Os conjuntos de dados foram submetidos a três classificações: total, amostra de treino (80% do conjunto total de dados) e amostra de teste (20% do conjunto total de dados), como é padrão em rotinas de aprendizado de máquina. Os resultados de desempenho foram aferidos quanto à consistência da taxa de precisão, coeficiente e parâmetro SAR em função da dimensão da amostra, na direção sugerida por Nicenboim *et al.* (2018) para distinguir amostra exploratória e amostra confirmatória.

A segunda rotina considerou o efeito do tipo de condições inseridas no modelo no resultado da classificação. Foram realizadas quatro classificações (com todas as variáveis; somente com variáveis estruturais, somente com variáveis prosódicas, somente com variáveis subjetivas), e igualmente os resultados de desempenho foram aferidos quanto à consistência da taxa de precisão,

coeficiente e parâmetro SAR em função do tipo das condições inseridas no modelo.

Para todas as rotinas, a classificação foi gerada pelo algoritmo classificador, com os parâmetros \bar{b} e C.45, e depois avaliada pela função `evaluate_Weka_classifier`, que realiza a validação cruzada da classificação implementada por 10 vezes, em arranjos aleatórios. Os resultados da validação cruzada são a base para a avaliação efeito do tamanho da amostra e do tipo de variáveis no desempenho da classificação.

As medidas de desempenho são a taxa de precisão da classificação (percentual das classificações corretas na validação cruzada), a estatística Kappa (κ) e o parâmetro da equação SAR (SAR), cujo resultado é um número entre 0 e 1.

A interpretação dos resultados de estatística para a concordância de classificação segue o que é consagrado na literatura (FLEISS; COHEN, 1973), em que valores abaixo de zero significam concordância menor que uma associação por coincidência; entre 0 e 0.20 a concordância não é significativa; entre 0.21 e 0.40 a concordância é baixa; entre 0.41 e 0.50 a concordância é moderada; entre 0.61 e 0.80 a concordância é substancial, e acima de 0.81 a concordância é ótima.

O cálculo da estatística considera o grau de concordância na precisão da classificação cruzada nas 10 reamostragens realizadas, enquanto a precisão é a média das classificações corretas. Quando a precisão se aproximam, sugerem a acurácia do aprendizado do algoritmo de classificação; quando há divergência entre estes valores, é possível que ocorrências previstas para uma dada classificação não foram realizadas, o que pode ser observado na matriz de confusão. Não há valores de referência para o parâmetro SAR; neste estudo consideramos a variância entre 0 e 1.

4 Resultados

4.1 Diminutivos

O conjunto de dados relativos ao diminutivo, com 241 observações, conta com dois objetivos de classificação: i) se o diminutivo é lexical ou afetivo; ii) se o diminutivo afetivo tem apreciação positiva ou negativa. Para implementar a rotina de classificação, os dados foram subdivididos em dois conjuntos, um com a totalidade de observações e a decisão de lexical ou afetivo ($n = 241$), com 0% de ocorrências classificadas como lexical, e 0% como afetivos; e outro conjunto apenas com os valores afetivos ($n = 200$), com 65.5% de ocorrências de decisão de valor positivo e 34.5% de ocorrências de decisão de valor negativo.

4.1.1 Tamanho da amostra

Para a decisão quanto ao tipo de diminutivo, se afetivo ou lexicalizado, no conjunto total dos dados, a taxa de precisão da classificação cruzada é de 94.19%, com $\kappa = 0.79$ e $SAR = 0.932$. No conjunto de treino (exploratório), com 80% da amostra, os resultados se mantêm estáveis em relação ao conjunto total: a taxa de precisão é de 93.23%, com $\kappa = 0.759$ e $SAR = 0.924$. No conjunto de teste (confirmatório), com 20% do conjunto de dados, a taxa de precisão é de 97.96%, com $\kappa = 0.921$ e $SAR = 0.985$. Os resultados mostram que o aprendizado do algoritmo leva a uma classificação que distingue diminutivos entre lexicalizados e afetivos que foi bem-sucedida, mas com diferença entre a etapa exploratória e confirmatória no grau de concordância.

	Precisão (%)	κ	SAR
total	94.19	0.790	0.932
treino	93.23	0.759	0.924
teste	97.96	0.921	0.985

Table 2. TABELA 1 - Comparação dos parâmetros de classificação quanto aos conjuntos de dados da classificação dos diminutivos entre lexicalizados e afetivos.

Já para a decisão de classificação entre apreciação positiva e negativa, no conjunto total de diminutivos afetivos, a taxa de precisão é de 75%, com $\kappa = 0.439$ e SAR = 0.842. A amostra treino tem 69.38% de precisão, $\kappa = 0.349$ e SAR = 0.809. Na amostra teste, a taxa de precisão e o coeficiente são os mesmos que os do conjunto de treino, variando apenas o parâmetro SAR, significativamente menor no conjunto de teste.

	Precisão (%)	K	SAR
total	75.00	0.439	0.842
treino	69.38	0.349	0.809
teste	69.38	0.349	0.709

Table 3. TABELA 2 - Comparação dos parâmetros de classificação quanto aos conjuntos de dados na classificação dos diminutivos afetivos entre positivos e negativos.

Considerando a evidência anterior, este resultado sugere que, para a divisão entre conjunto de treino (exploratório) e teste (confirmatório), o conjunto total de dados precisa ser robusto para barrar geração de entropia na classificação (no caso, passou de moderada a baixa), seja positiva ou negativa.

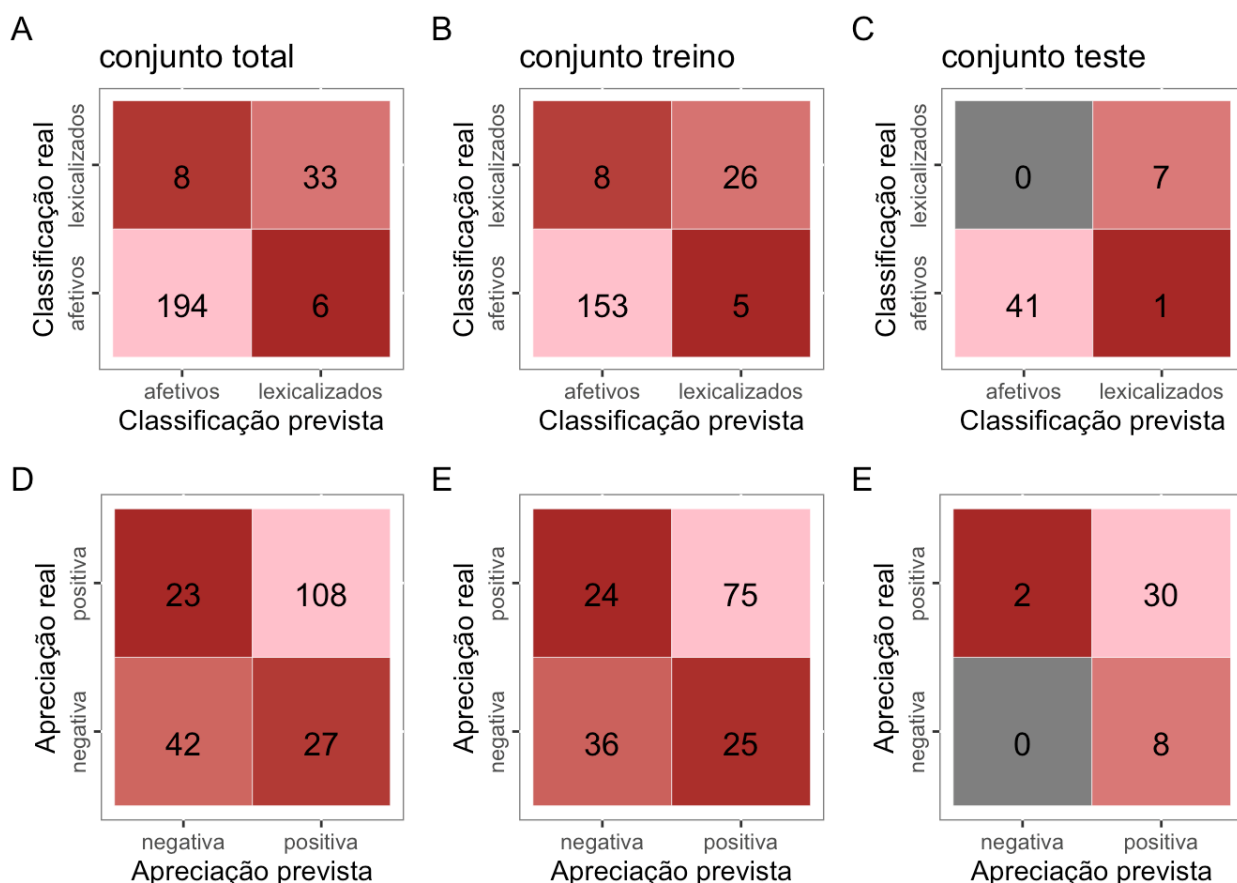


Figure 4. FIGURA 4 - Matrizes de confusão das diferentes classificações realizadas com os dados dos diminutivos quando foram classificados como diminutivos lexicais ou afetivos (em A, B e C) ou quando o diminutivo afetivo tem apreciação positiva ou negativa (em D, E e F). Em A e D, com o conjunto completo dos dados. Em B e E, com o conjunto de treino, e em C e F, com o conjunto de teste.

A matriz de confusão das classificações permite visualizar qual dos dois níveis das variáveis apresenta maior problema no aprendizado do algoritmo. No conjunto de teste, enquanto na decisão de classificar os diminutivos como lexicalizados ou afetivos não houve ocorrência de diminutivo lexicalizado classificado como afetivo, na decisão de classificar os diminutivos afetivos como negativos ou positivos, não houve ocorrência de diminutivo afetivo negativo classificado como negativo, o que se manifesta pelo coeficiente baixo. Novamente, este pode ser um efeito do tamanho da amostra no aprendizado do algoritmo.

4.1.2 Tipo de condições

Para verificar o efeito dos tipos das condições controladas no aprendizado do algoritmo para a classificação dos diminutivos entre lexicalizados e afetivos, foram realizadas classificações considerando os tipos de variáveis: estruturais, prosódicas e subjetivas.

A taxa de precisão para as variáveis estruturais é de 94.61%, com $\kappa = 0.803$ e SAR = 0.931. Já o comportamento das variáveis prosódicas e subjetivas é o mesmo: a taxa de precisão de 82.99%, $\kappa = 0$ e SAR = 0.772.

	Precisão (%)	κ	SAR
todas as variáveis	94.19	0.790	0.932
estruturais	94.61	0.803	0.931
subjetivas	82.99	0.000	0.772
prosódicas	82.99	0.000	0.772

Table 4. TABELA 3 - Comparação dos parâmetros de classificação quanto aos tipos de variáveis na classificação dos diminutivos entre lexicalizados e afetivos.

Comparando com os parâmetros do conjunto total, observamos que são as variáveis estruturais as responsáveis pela acurácia do aprendizado do algoritmo de classificação para a decisão entre diminutivos lexicalizados ou afetivos.

	Precisão (%)	κ	SAR
todas as variáveis	75.0	0.439	0.842
estruturais	67.5	0.091	0.727
subjetivas	64.5	0.154	0.775
prosódicas	65.5	0.000	0.713

Table 5. TABELA 4 - Comparação dos parâmetros de classificação quanto aos tipos de variáveis na classificação dos diminutivos afetivos entre positivos e negativos.

Quanto à apreciação, se positiva ou negativa, são as variáveis subjetivas que apresentam maior contribuição para o aprendizado do algoritmo de classificação: embora com 64.5% de precisão, o grau de concordância é de $\kappa = 0.154$, com parâmetro SAR = 0.775. As variáveis estruturais na classificação dos diminutivos afetivos têm taxa de precisão equivalente, mas $\kappa = 0.091$, não significativo, e SAR = 0.727. As variáveis prosódicas têm contribuição próxima do acaso ($\kappa = 0$).

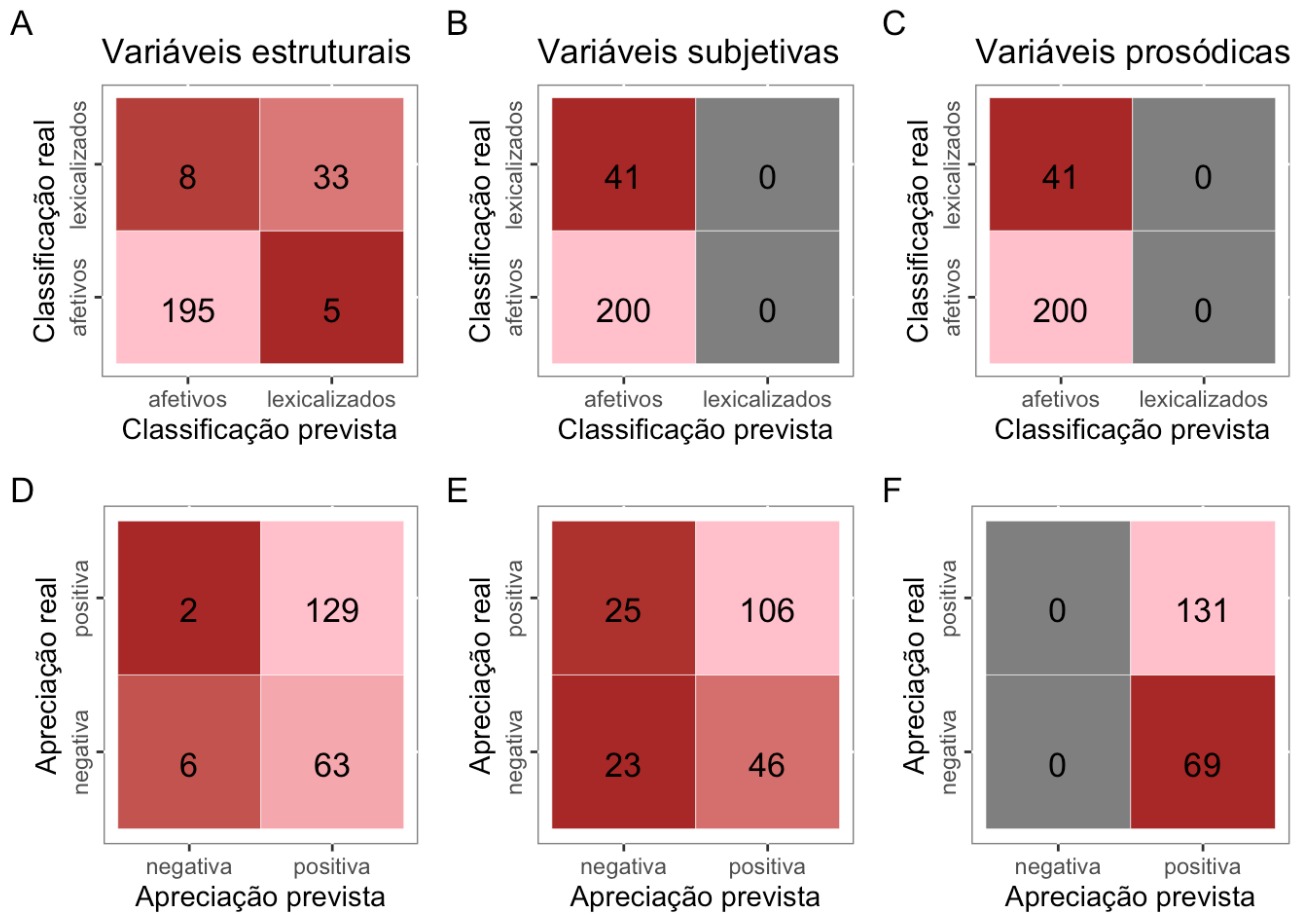


Figure 5. Figura 5 – Matrizes de confusão das diferentes classificações realizadas com os dados dos diminutivos quando foram classificados como diminutivos lexicais ou afetivos (em **A**, **B** e **C**) ou quando o diminutivo afetivo tem apreciação positiva ou negativa (em **D**, **E** e **F**). Em **A** e **D**, com as variáveis estruturais. Em **B** e **E**, com as variáveis subjetivas, e em **C** e **F**, com as variáveis prosódicas.

Observando a matriz de confusão da classificação dos diminutivos quanto à contribuição das variáveis, identificamos o efeito não significativo das variáveis subjetivas e prosódicas na decisão entre lexicalizados e afetivos: nenhum diminutivo lexicalizado previsto foi classificado pelo algoritmo na classificação cruzada, o que explica o coeficiente perto do aleatório. Na decisão entre positiva e negativa para a apreciação dos diminutivos afetivos, 63 ocorrências previstas como afetivas negativas foram classificadas como positivas, o que também explica o coeficiente perto do aleatório. Já quanto às variáveis prosódicas, não houve ocorrência prevista como negativa classificada como negativa.

4.2. Modalizador

O conjunto de dados relativos ao modalizador (*eu acho que*) conta com 1038 observações, das quais 78.52% são de valor de certeza, 9.25% são de dúvida e 12.24% são de incerteza.

4.2.1 Tamanho da amostra

Embora a taxa de precisão seja muito próxima nos três conjuntos de dados de valores do modalizador (*eu acho que*), (78.52% ~ 78.85%), o grau de concordância é próximo do acaso, com = 0 para o conjunto total e = 0 para o conjunto de treino. Já no conjunto de testes (20% da amostra, n = 208, o grau de concordância é baixo ($\kappa = 0.243$). Do mesmo modo que no conjunto dos diminutivos

afetivos, encontramos diferenças entre a análise exploratória e a análise confirmatória, reforçando a importância de amostras robustas para a classificação dos dados.

	Precisão (%)	K	SAR
total	78.52	0.000	0.758
treino	78.67	0.000	0.759
teste	78.85	0.243	0.789

Table 6. TABELA 5 - Comparação dos parâmetros de classificação quanto aos conjuntos de dados da classificação dos valores do modalizador (*eu acho que*).

Se nos diminutivos afetivos a amostra de teste era muito reduzida, o mesmo não se pode dizer da amostra de teste para (*eu acho que*), o que sugere que a entropia esteja associada às condições controladas.

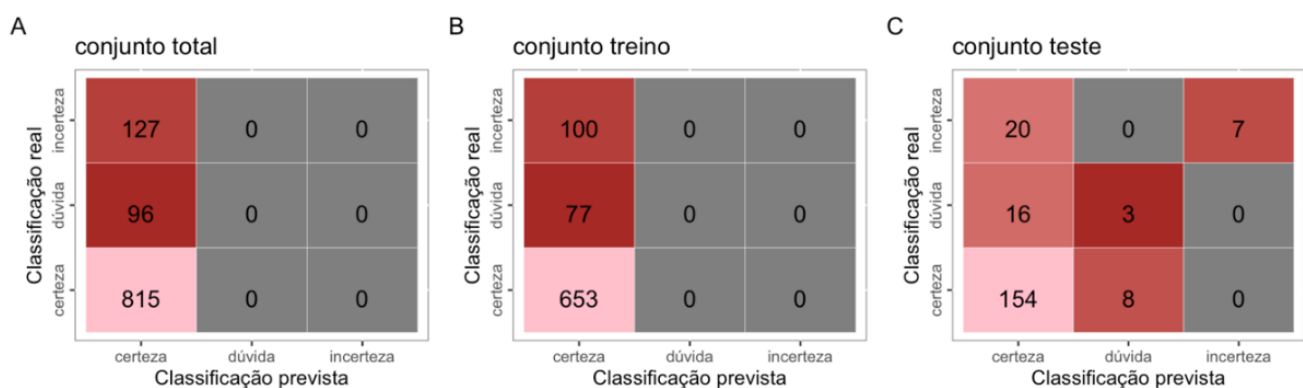


Figure 6. FIGURA 6 - Matrizes de confusão das diferentes classificações realizadas com os dados de (*eu acho que*) quando foram classificados como modalizador de certeza, dúvida ou incerteza com o conjunto total de dados (A), com o conjunto de treino (B), e com o conjunto de teste (C).

Na matriz de confusão, observamos que o algoritmo não consegue classificar ocorrências que não sejam de certeza; somente na amostra de treino, que é composta por 20% do total, a classificação prevista e realizada para os valores de dúvida e incerteza apareceram. Comparando com os resultados obtidos para os diminutivos, constatamos que amostras reduzidas podem interferir no aprendizado do classificador, afetando positiva ou negativamente a sua acurácia.

4.2.2 Tipo de condições

Na atribuição de valores do modalizador (*eu acho que*), o tipo das variáveis não interfere no algoritmo de classificação: a taxa de concordância é de ~ 78%, e o grau de concordância é perto do acaso, com próximo de zero ou negativo.⁶

	Precisão (%)	K	SAR
todas as variáveis	78.52	0.000	0.758
estruturais	78.13	-0.006	0.774
subjetivas	78.03	0.016	0.773
prosódicas	78.32	-0.003	0.775

Table 7. TABELA 6 - Comparação dos parâmetros de classificação quanto aos tipos de variáveis na classificação dos valores do modalizador (*eu acho que*).

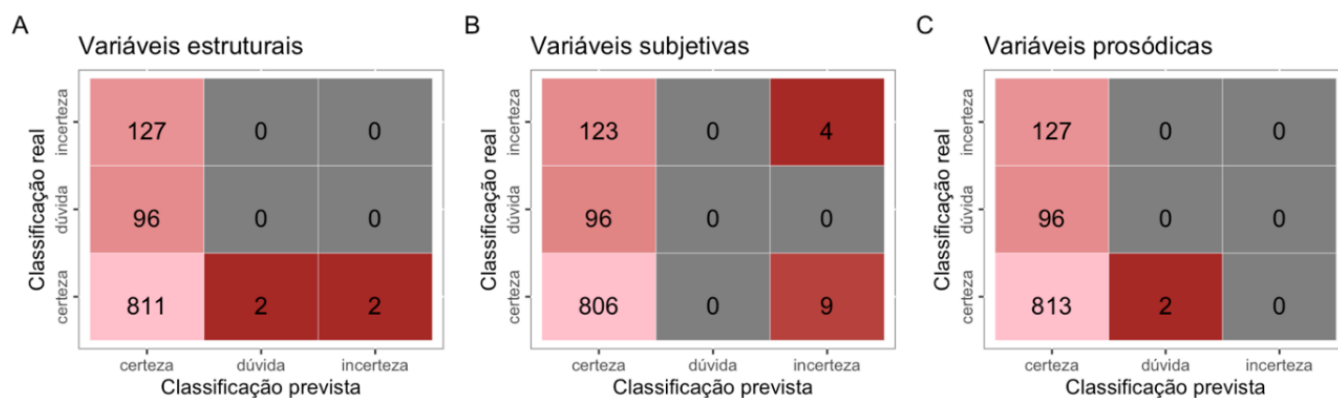


Figure 7. FIGURA 7 - Matrizes de confusão das diferentes classificações realizadas com os dados de (*eu acho* que quando foram classificados como certeza, dúvida ou incerteza com as variáveis estruturais (A), subjetivas (B), e prosódicas (C).

Embora o conjunto de dados de modalizador (*eu acho* que seja substancialmente mais robusto do que o de diminutivos, as variáveis controladas não permitem uma classificação com acurácia, sugerindo que o analista se valha de outras pistas para a atribuição dos valores (e que não foram incluídas no modelo), seguindo o que apontam Kononenko e Bratko (1991).

5. Discussão

Após a utilização de metodologias provenientes do aprendizado de máquinas para avaliar a consistência dos critérios adotados ao analisar dois conjuntos de dados, pode-se afirmar que o tamanho da amostra interfere no aprendizado do algoritmo de classificação: os conjuntos de teste apresentam comportamento diferente dos conjuntos de treino, em ambos os fenômenos, interferindo no algoritmo de classificação positiva ou negativamente em relação aos padrões anteriores.

Variáveis estruturais contribuem para a acurácia do algoritmo de classificação para diferenciar entre diminutivos lexicalizados e afetivos, mas entre diminutivos afetivos, variáveis subjetivas são medianamente acuradas para diferenciar entre apreciação negativa e positiva. Variáveis prosódicas não contribuem para o algoritmo de classificação nem de diminutivos nem de valores do modalizador (*eu acho* que). As matrizes de confusão permitem visualizar os contextos de classificação previstos e esperados; os zeros em contextos previstos não realizados sinalizam quais são os níveis da variável que é objeto da classificação que são mais responsáveis pela entropia do modelo; no caso, diminutivos afetivos negativos, e modalização de dúvida e de incerteza de (*eu acho* que).

A separação entre análise exploratória e análise confirmatória sugerida por Nicenboim *et al.* (2018) para avaliar a consistência de resultados e acurácia de modelo requer, como evidenciamos, que a amostra inicial seja robusta o suficiente para permitir rearranjos na classificação cruzada (10 seleções aleatórias dentro do conjunto de dados); no caso dos diminutivos afetivos, cujo conjunto total é composto de 41 ocorrências, o conjunto de teste (20% da amostra) foi menor do que o número total de seleções, o que pode ter interferido no resultado.

Apesar de os fatores prosódicos fornecerem medidas precisas, e que se mostraram estatisticamente significativas na distinção entre funções, tanto dos diminutivos quanto do modalizador (ANTUNES; AUBERGÉ, 2015; FREITAG; CARDOSO; PINHEIRO, 2020; VALDERRAMA; PÉREZ; FUENTES, 2017), o algoritmo de classificação não considerou estas condições, evidenciando a necessidade de um refinamento do controle ou mudança na medida.

A adoção de três medidas para o desempenho do algoritmo classificador chama a atenção para os casos de não convergência: embora a taxa de precisão seja relativamente alta em todos os modelos, o grau de concordância variou significativamente, desde a aleatoriedade até a concordância ótima. O parâmetro SAR (JAPKOWICZ; SHAH, 2011) apresentou um escopo de variação estável e restrito, não tão saliente quanto a comparação entre a taxa de precisão e o coeficiente κ . Assim como a diversidade de técnicas para a classificação de conjuntos de dados, a diversidade de medidas também é importante para a ampliação do poder de generalização dos valores de funções para um conjunto maior de usos da língua.

Quanto à replicabilidade, os resultados da classificação do algoritmo J48 sugerem que uma replicação de uma classificação de diminutivos entre lexicalizados e afetivos, a partir das condições controladas neste modelo, seria bem-sucedida; já para *(eu) acho que*, as condições incluídas no modelo não permitem que uma replicação tenha resultados convergentes. Destaque-se que dentre as condições que mais contribuíram para o aprendizado do algoritmo de classificação foram as variáveis subjetivas, que são também de natureza inferencial do analista, levando a um *loop* de inferências cuja consequência é a impossibilidade de replicação.

Além do aprendizado de um algoritmo de classificação, como propomos, a ampliação de analistas para uma mesma função (seja por meio de tarefa experimental, seja por meio de classificação de juízes especialistas) pode aumentar o poder explanatório das classificações de funções, e permitir a reprodutibilidade das pesquisas.

Com este estudo, contribuimos para o aprimoramento de um *modus operandi* inerente à linguística, apresentando uma solução para validar a intuição e sensibilidade do analista que codifica a regra, traduzida em termos de valores de uma função. A testagem dos critérios adotados por um analista por meio de aprendizagem de máquina permite avaliar a sua consistência e potencialidade de generalização. Do mesmo modo, permite identificar quando as condições de um modelo não são as responsáveis pela atribuição do julgamento.

References

1. ALVES E. O diminutivo no português do Brasil: funcionalidade e tipologia. *Estudos Linguísticos*. 2006.
2. ANTUNES L. B, AUBERGÉ V. Análise Prosódica da certeza e da incerteza em fala espontânea e atuada. *Revista Diadorim*. 2015.
3. BARBOSA M. F, FREITAS M. A. A distribuição e o uso dos diminutivos -inho e -zinho no português brasileiro: uma abordagem pela fonologia de uso. *Revista do DLCV*. 2014.
4. BEREZ-KROEKER A. L, et al.. Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics*. 2018.
5. BOERSMA P, VAN HEUVEN V. Speak and unSpeak with PRAAT. *Glott International*. 2001.
6. CARDOSO P. B.. Efeitos paralinguísticos na inferência de sentidos indicados por *(eu) acho que* em entrevistas sociolinguísticas. Dissertação (Mestrado em Letras) - Universidade Federal de Sergipe; Pós-Graduação em Letras, 2021.
7. CARUANA R, et al. Ensemble selection from libraries of models. *Anais....* 2004.
8. CARVALHO F, GUEDES G. P. Avaliação das categorias afetivas do dicionário PB-LIWC2015. *Anais....* 2019.
9. CASSEB-GALVÃO V. C. O achar no português do Brasil: um caso de gramaticalização. *Sínteses*. 2000.
10. COSTA I. Processos morfofonológicos na morfologia derivacional. Unicamp; 1993.
11. CROFT W. Typology and universals. 2002.
12. FLEISS J. L, COHEN J. . The equivalence of weighted kappa and the intraclass correlation

- coefficient as measures of reliability. *Educational and psychological measurement*. 1973.
13. FREITAG R. O uso de Tá? e Certo? na fala de Santa Catarina. *Working Papers em Linguística*. 2001.
 14. FREITAG R. M. K. “Mudar para variar”, “variar para mudar” –tratando da variação e mudança de acho (que) e parece (que) parentéticos epistêmicos na fala de Florianópolis. *Fórum Linguístico*. 2007.
 15. FREITAG R. M. K. Variação em categorias verbais: correlações entre forma e função. *Estudos Linguísticos*. 2011.
 16. FREITAG R. M. K. Dissecando a entrevista sociolinguística: estilo, sequência discursiva e tópico. *Variação estilística: reflexões teórico-metodológicas e propostas de análise*. 2014.
 17. FREITAG R. M. K. Kappa statistic for judgment agreement in Sociolinguistics/Estatística Kappa para concordância de julgamento em Sociolinguística. *Revista de Estudos da Linguagem*. 2019.
 18. FREITAG R. M. K, CARDOSO P. B. , PINHEIRO B. F. M. Em busca de pistas prosódicas para a desambiguação de sentidos pragmáticos: estudo piloto com diminutivos e modalizadores parentéticos epistêmicos em português brasileiro. *Caderno de resumos I Encontro Online de Fonética e Fonologia (EN-OFF)*. 2020.
 19. FREITAG R. M. K, PINHEIRO B. F. M. Modelo de árvore de inferência condicional para explicar usos linguísticos variáveis. EDUNEB; 2020.
 20. FREITAG R. M. K, SILVA R. B, EVANGELISTA F. R. Marcadores discursivos interacionais: diferentes metodologias, diferentes resultados. *Diacrítica*. 2017.
 21. FURTADO DA CUNHA M. A, OLIVEIRA M. R, VOTRE S. A interação sincronia/diacronia no estudo da sintaxe. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*. 1999.
 22. GARNER S. R. Weka: The waikato environment for knowledge analysis. *Anais....* 1995.
 23. GIVON T. Typology and functional domains. *Studies in Language*. 1981.
 24. GOODFELLOW I, BENGIO Y, COURVILLE A. Machine learning basics. *Deep learning*. 2016.
 25. HASAN M, et al. A study of the effectiveness of machine learning methods for classification of clinical interview fragments into a large number of categories. *Journal of biomedical informatics*. 2016.
 26. HEINE B, CLAUDI U, HÜNNEMEYER F. From Cognition to Grammar — Evidence from African Languages. 1991.
 27. HORNIK K, BUCHTA C, ZEILEIS A. Open-source machine learning: R meets Weka. *Computational Statistics*. 2009.
 28. JAPKOWICZ N, SHAH M. Evaluating learning algorithms: a classification perspective. 2011.
 29. KONONENKO I, BRATKO I. Information-based evaluation criterion for classifier’s performance. *Machine learning*. 1991.
 30. MAIRESSE F, et al. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research*. 2007.
 31. NICENBOIM B, et al.. Exploratory and confirmatory analyses in sentence processing: A case study of number interference in German. *Cognitive science*. 2018.
 32. NICHOLS J. Functional theories of grammar. *Annual review of Anthropology*. 1984.
 33. OLIVEIRA B. A prosódia na expressão das atitudes de dúvida, incerteza e incredulidade no português brasileiro. Dissertação (Mestrado em Linguística) — Faculdade de Letras, Universidade Federal do Rio de Janeiro, 2011.
 34. PEELS R, BOUTER L. The possibility and desirability of replication in the humanities. *Palgrave Communications*. 2018.
 35. PINHEIRO B. F. M. Pistas linguísticas e paralinguísticas para os sentidos dos diminutivos. Dissertação (Mestrado em Letras) - Universidade Federal de Sergipe, Pós-Graduação em Letras, 2021.
 36. QUINLAN R. C4.5: Programs for Machine Learning. 1993.
 37. ROCHA R. M, VICENTE H S. Diminutivos: uma análise comparativa entre o português e o inglês. *Signum: Estudos da Linguagem*. 2016.
 38. SANTANA M. M. S. O sufixo diminutivo em português: forma, funcionamento e significação-do século XIII ao XX. Tese (Doutorado) - Universidade de São Paulo, 2017.

39. SANTANA R. Tipos de Tipo em uma comunidade de práticas universitária. Dissertação (Mestrado em Letras) - Universidade Federal de Sergipe; Pós-Graduação em Letras, 2019.
40. SIMAKI V, MPORAS I, MEGALOOIKONOMOU V. Age identification of twitter users: Classification methods and sociolinguistic analysis. 2016.
41. STIGLIC G, et al. Comprehensive decision tree models in bioinformatics. *PloS one*. 2012.
42. TEIXEIRA T.. A forma e o uso dos sufixos -inho e -zinho em variedades do sul do Brasil. Tese (Doutorado) -Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 2008.
43. THOMPSON S. A, MULAC A. A quantitative perspective on the grammaticization of epistemic parentheticals in English. 1991.
44. TORRES-CACOULOS R. From lexical to grammatical to social meaning. *Language in society*. 2001.
45. TRAUGOTT E. C, DASHER R. B. Regularity in semantic change. 2004.
46. VALDERRAMA R, PÉREZ H. E, FUENTES D. Incidencia del valor pragmático en la duración de la vocal acentuada del sufijo diminutivo-ito en el español de Chile. *Boletín de filología*. 2017.
47. VILLALVA A. Estruturas Morfológicas: unidades e hierarquias nas palavras do português. 2000.
48. WALKER J. A. Modelling Socio-grammatical Variation. *Advancing Socio-grammatical Variation and Change: In Honour of Jenny Cheshire*. 2020.
49. WIELING M, RAWEE J, NOORD G. V. Reproducibility in computational linguistics: Are we willing to share?. *Computational Linguistics*. 2018.