

Avaliação Automática da Complexidade de Sentenças do Português Brasileiro para o Domínio Rural

Sidney E. Leal¹, Vanessa M. A. Magalhães², Magali S. Duran¹, Sandra M. Aluísio¹

¹Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo (USP)
Caixa Postal 668 – 13560-970 – São Carlos – SP

²Núcleo de Gestão da Informação e Conhecimento, Embrapa Gado de Leite
Juiz de Fora - MG

¹sidleal@gmail.com, magali.duran@uol.com.br, sandra@icmc.usp.br

²vanessa.magalhaes@embrapa.br

Abstract. Low literacy is a common problem in the Brazilian dairy sector that may undermine productivity. Hence the importance of simplifying newsletters, technical texts and instructions addressed to this public. The task of automatic evaluation of sentential complexity is new to Portuguese and allows us, for example, to identify which sentences in a text should be simplified. This paper presents a 3-step method for this task, using classical machine learning with MLP neural networks for ranking and regression. The model was trained in a public corpus of sentences collected from journalistic texts and its generalization to other scenarios was evaluated for the rural domain. We obtained accuracy of 87.80% in the ranking, root-mean-square error (RMSE) of 0.06 in the regressor and F-measure of 88.4% in the robustness test.

Resumo. A maioria dos produtores de leite possuem baixo letramento, o que prejudica seu acesso às tecnologias com vistas à melhoria das condições de trabalho, bem como ao aumento na produção e na renda. Esse é o motivo da importância de simplificar informativos e textos técnicos dirigidos a esse público. A tarefa de avaliação automática da complexidade sentencial é nova para o português e permite identificar, por exemplo, quais sentenças em um texto devem ser alvo de simplificação. Este artigo apresenta um método de 3 passos para essa tarefa, utilizando redes neurais do tipo MLP para ranqueamento e regressão. O modelo foi treinado em um córpus público de sentenças do gênero jornalístico e sua generalização para outros cenários foi avaliada para o domínio rural. Foram obtidas uma acurácia de 87,80% no ranqueamento, raiz do erro quadrático médio (ou RMSE, em inglês) de 0,06 no regressor e F-measure de 88,4% no teste de robustez.

1. Introdução

Segundo o INAF Brasil (Indicador de Alfabetismo Funcional), apenas um em cada dez brasileiros adultos é considerado letrado de forma proficiente [IPM 2018]. Esse indicador é bastante alarmante e explicita um dos grandes desafios brasileiros: o acesso à evolução econômica e tecnológica pela população. Percebe-se que a situação é ainda mais crítica quando isolamos certos setores da economia, como o da agropecuária, em que apenas 1% dos entrevistados foram considerados proficientes. Isso significa que a quase totalidade dos produtores rurais pode não ser capaz de usufruir das novas tecnologias desenvolvidas

e publicadas por entidades de pesquisa. A falta de acesso ao conhecimento prejudica bastante esse setor, um dos mais importantes do Brasil, que é responsável por 23% do Produto Interno Bruto (PIB)¹ e 40% da renda da população economicamente ativa².

Uma alternativa viável na atualidade é simplificar essas publicações, como fez a Embrapa Gado de Leite, no projeto APP@Rural [Magalhães et al. 2017], que simplificou os comunicados técnicos e partes do Manual de Bovinocultura de Leite, tornando-os mais acessíveis aos produtores, estudantes e extensionistas. A Embrapa utiliza os métodos de classificação textual e de simplificação lexical e sintática do projeto PorSimples (Simplificação Textual do Português para Inclusão e Acessibilidade Digital) [Aluisio and Gasperin 2010], com adaptação da simplificação lexical para atender aos termos técnicos do domínio rural. O PorSimples teve como objetivo promover o acesso a textos em português brasileiro (PB) por analfabetos funcionais e crianças ou adultos em fase de alfabetização e criou os modelos automáticos com base em textos jornalísticos.

O trabalho presente propõe uma evolução nesses métodos ao criar um método para indicar automaticamente as sentenças alvos de simplificação, permitindo a sua classificação nos quatro níveis indicados pelo relatório de 2018 do INAF: Proficiente, Intermediário, Elementar e Rudimentar.

Na Seção 2, são apresentados a tarefa e os principais trabalhos da área de complexidade sentencial. Na Seção 3, é apresentado o método de avaliação da complexidade proposto neste trabalho, juntamente com o córpus de sentenças alinhadas PorSimplesSENT, um resumo das *features* e o modelo de aprendizado de máquina escolhido. Finalmente, a Seção 4 mostra o teste de robustez feito para o modelo treinado com textos jornalísticos e avaliado em textos produzidos pela Embrapa Gado de Leite, chamados aqui de textos do domínio rural.

2. Avaliação da Complexidade Sentencial

A inteligibilidade³ de um texto, do inglês *text readability* é, segundo [Dubay 2007, pg.6], *a facilidade de leitura de um texto criada pela escolha de conteúdo, estilo, estruturação e organização que atende ao conhecimento prévio, habilidade de leitura, interesse e motivação da audiência*. As primeiras fórmulas de inteligibilidade foram criadas há quase um século, na década de 1920, nos Estados Unidos e consideravam que a complexidade poderia ser inferida por métricas de palavras e sentenças, baseadas na frequência e tamanho (quantidade de letras) das palavras e na média da quantidade de palavras por sentença. Desde então, a Inteligibilidade Textual tornou-se uma grande área de pesquisa multidisciplinar, com uma vasta bibliografia, e ganhou novas abordagens neste século com o uso de métodos de PLN e AM.

Tradicionalmente, a tarefa tem sido aplicada no nível textual, atribuindo uma nota (ou nível de *ranking*, de proficiência) para um documento inteiro. Porém, em um documento classificado como simples, podem ocorrer sentenças complexas, assim como existem sentenças simples em um documento complexo. Uma sentença é uma unidade

¹<http://www.agricultura.gov.br/noticias/agropecuaria-puxa-o-pib-de-2017>

²<http://www.mda.gov.br/sitemda/noticias/agricultura-familiar-do-brasil-%C3%A9-8%C2%AA-maior-produtora-de-alimentos-do-mundo>

³Neste trabalho, usamos os termos complexidade e inteligibilidade (seja no nível sentencial ou textual) como sinônimos.

importante que traz, na maioria das vezes, informação suficiente para inferência e análise da sua complexidade. Embora seja possível usar a mesma abordagem de avaliação da complexidade dos textos para o nível das sentenças, [Dell'Orletta et al. 2014] demonstraram que é necessário um número maior de *features* para a segunda tarefa.

A Tabela 1 mostra: a) uma simplificação por meio da substituição lexical, em que um termo técnico é substituído por outro mais frequente; b) uma sentença simplificada no nível sintático por meio da sua divisão em duas sentenças e c) um caso de elaboração textual, incluindo uma breve explicação de um termo técnico.

Tabela 1. Exemplos de sentenças simplificadas

a) Lexicalmente	Original Simplificada	Se acentuada e prolongada, a hipertermia pode causar a morte do animal. Se acentuada e prolongada, a febre pode causar a morte do animal.
b) Sintaticamente	Original Simplificada	O uso de forragem conservada, cujas formas mais comuns são: ensilagem e fenação, é uma solução para alimentar o rebanho. O uso de forragem conservada é uma solução para alimentar o rebanho. As formas mais comuns para conservar forragens são: ensilagem e fenação.
c) Elaboração textual	Original Simplificada	A ensilagem é o processo de conservação do alimento por fermentação anaeróbia. A ensilagem é o processo de conservação do alimento por fermentação anaeróbia (sem a presença de ar).

A avaliação do nível de inteligibilidade de sentenças é uma tarefa de pesquisa recente e visa analisar e avaliar individualmente as sentenças de um texto, permitindo uma informação mais acurada dos pontos complexos de um texto. Essa abordagem é relevante, pois, como afirmam [Dell'Orletta et al. 2014] as abordagens de classificação de inteligibilidade que levam em consideração os textos inteiros não trazem grandes vantagens para a posterior aplicação de métodos automáticos de simplificação. Além disso, considerar como complexas todas as sentenças de um texto classificado como complexo, pode prejudicar o treinamento dos métodos, principalmente quando essas sentenças são utilizadas para avaliar a tarefa de predição da complexidade sentencial. Isso foi demonstrado por [Vajjala and Meurers 2014] durante a investigação dos motivos da baixa acurácia que obtiveram ao utilizar o córpus Wikipedia-SimpleWikipedia (sem alinhamento sentencial).

O primeiro trabalho a considerar a tarefa de complexidade especificamente para o nível sentencial foi [Dell'Orletta et al. 2011], comparando a sua dificuldade em relação ao nível textual. Porém, a definição da forma de avaliação da tarefa só foi consolidada por [Vajjala and Meurers 2016] e permitiu que os trabalhos posteriores aperfeiçoassem os resultados comparativamente. [Ambati et al. 2016] conseguiram melhorar significativamente os resultados utilizando um parser do tipo *Combinatory Categorial Grammar* (CCG), e [Gonzalez-Garduño and Søgaard 2018] chegaram no estado da arte para o inglês, utilizando métricas de rastreamento ocular aliadas às linguísticas e psico-linguísticas. [Howcroft and Demberg 2017] e [Singh et al. 2016] também publicaram trabalhos explorando a tarefa com novas métricas; o primeiro trabalho exclusivamente

com métricas psicolinguísticas e o segundo com métricas de rastreamento ocular. Uma comparação dos resultados obtidos para a tarefa de avaliação da complexidade sentencial na língua inglesa é mostrada na Tabela 2.

Tabela 2. Avaliação da tarefa na Wikipedia-SimpleWikipedia

Trabalho	Método	Acurácia
Flesch-Kincaid	Método Clássico	72,30
[Vajjala and Meurers 2016]	RankSVM	74,58
[Ambati et al. 2016]	SMO/Incr CCG	78,87
[Singh et al. 2016]	Regressão Log.	75,21
[Howcroft and Demberg 2017]	RankAsClass.	73,22
[Gonzalez-Garduño and Søgaard 2018]	MultiTask MLP	86,62

Mais recentemente, [Stajner et al. 2017] e [Scarton et al. 2018] contribuíram para a tarefa, avaliando a complexidade com o apoio do córpus Newsela (que possui 550 mil sentenças, três vezes maior que o Wikipedia-SimpleWikipedia) e [Bosco et al. 2018] obtiveram bons resultados para o italiano, utilizando Redes Neurais Recorrentes do tipo *Long Short Term Memory* (LSTM). Finalmente, [Brunato et al. 2018] contribuíram com um trabalho sobre a percepção da complexidade e concordância entre anotadores, enquanto [Timm 2018] investigou simplificações sentenciais automáticas, utilizando rastreamento ocular. Para o português (até onde foi possível verificar), foi encontrado apenas o trabalho de [Leal et al. 2018], em que foi publicado o córpus PorSimpleSent, com foco nesta tarefa.

3. Método de Avaliação da Complexidade Sentencial para o Português

3.1. Córpus

O PorSimpleSent [Leal et al. 2018] é um córpus de sentenças alinhadas, disponível publicamente, que foi compilado a partir do PorSimple [Caseli et al. 2009] e organizado a partir do alinhamento sentencial dos textos, em três níveis: a) **Original**: Sentenças originais; b) **Simplificação Natural**: Textos simplificados de forma livre pelos anotadores e c) **Simplificação Forte**: Textos simplificados seguindo as regras do manual desenvolvido no projeto.

O córpus PorSimpleSent possui três versões, com diferentes abordagens para as sentenças que sofreram operação de divisão. O PSS1 repete a sentença original para cada sentença resultante da divisão; o PSS2 seleciona apenas a maior sentença resultante, que também tenha maior sobreposição de palavras, e o PSS3 contém apenas sentenças que não sofreram divisão, sendo, portanto, o menor dos três. Para este trabalho foi escolhida a versão **PSS2**, que possui 4.962 pares de sentenças, com alinhamentos Original-Natural, Natural-Forte e Original-Forte, obtida no formato TSV (*Tab Separated Values*)⁴.

3.2. Features

Este trabalho utiliza como *features* as métricas disponibilizadas pelas ferramentas Coh-Metrix-Port, Coh-Metrix-Dementia, LIWC, AIC e as métricas psicolinguísticas disponibilizadas por [dos Santos et al. 2017], que anotaram automaticamente um banco⁵

⁴<http://www.nilc.icmc.usp.br/nilc/index.php/tools-and-resources>

⁵A base está disponível em: <http://143.107.183.175:21380/portlex/index.php/en/component/content/article/2-uncategorised/23-psycholinguistic>

de 26.874 palavras do PB com Imageabilidade, Concretude, Familiaridade e Idade de Aquisição. O conjunto utilizado totaliza 189 features.

O Coh-Metrix-Port⁶ [Scarton and Aluísio 2010] é uma adaptação para o PB do Coh-Metrix, desenvolvida dentro do projeto PorSimples, e implementa 48 métricas, divididas nas categorias: contagens básicas, operadores lógicos, frequências, hiperônimos, tokens, constituintes, conectivos, ambiguidade, co-referência e anáforas. O Coh-Metrix-Dementia⁷ [Cunha 2015] é uma adaptação do Coh-Metrix-Port para análise automática de distúrbios de linguagem nas demências (como Doença de Alzheimer) ou no Comprometimento Cognitivo Leve (CCL). Ele adiciona 25 novas métricas às 48 do Coh-Metrix-Port, nas categorias: disfluências, análise de semântica latente, diversidade lexical, complexidade sintática e densidade semântica. LIWC (*Linguistic Inquiry and Word Count* - liwc.wengine.com) é uma ferramenta baseada em dicionários para análise dos vários componentes emocionais, cognitivos e linguísticos em amostras de textos, com categorias como: estatísticas comuns do texto, dimensão linguística, processos psicológicos, relatividade, assuntos pessoais e miscelânea, totalizando aproximadamente 100 métricas [Cunha 2015]. A tradução e adaptação do dicionário para o PB foi realizada em uma colaboração entre NILC, Checon Pesquisa e Unisinos no período de 2010 a 2012 e está disponível no site do projeto PortLex⁸. Também criada dentro do contexto do PorSimples [Maziero et al. 2008], a ferramenta AIC (Análise Automática de Inteligibilidade de Córpus) traz 39 métricas, com o principal diferencial de utilizar o analisador sintático PALAVRAS [Bick 2000] para o cálculo delas. Elas estão organizadas em seis classes: estatísticas do texto, voz passiva, características das orações, densidade, personalização e marcadores discursivos [Cunha 2015].

3.3. Modelo Proposto

O treinamento do modelo foi feito em três fases, ilustradas na Figura 1, na qual é feita uma analogia entre a complexidade e o espectro de cores (vermelho = mais complexo, azul = mais simples). Neste exemplo, uma sentença nunca vista antes (F3), de cor verde mais intensa que o verde da posição 4, tem o valor estimado de complexidade 4,5 (entre 4-verde e 5-amarelo).

Fase 1 - Pairwise Ranking

A primeira tarefa consistiu em treinar uma rede neural MLP (*Multi Layer Perceptron*) com 3 camadas, utilizando os 4.962 pares de sentenças do córpus PSS2 e todas as *features* disponíveis. Metade dos pares foi invertida, de forma a balancear melhor as duas classes: o lado complexo foi anotado com 1 e o lado simples com 0.

A camada de entrada da MLP contou com 378 neurônios (189 *features* x 2 sentenças); a camada oculta foi configurada com 30 neurônios, utilizando a função de ativação *ReLU*, e a saída possui apenas um neurônio, com a função *sigmoid*, predizendo 1 quando a sentença A é mais complexa que a B e 0 no inverso. O ranqueador conseguiu uma acurácia de **87,8%**, utilizando *10 fold cross validation*, um pouco superior ao estado da arte da tarefa para a língua inglesa no córpus Wikipedia-SimpleWikipedia. Utilizamos como *baseline* os resultados reportados por [Leal et al. 2018], cujo melhor modelo atin-

⁶<http://fw.nilc.icmc.usp.br:22680/>

⁷<http://fw.nilc.icmc.usp.br:22380/>

⁸<http://nilc.icmc.usp.br/portlex/index.php/pt/projetos/liwc>

Figura 1. Diagrama da sequência de passos do modelo

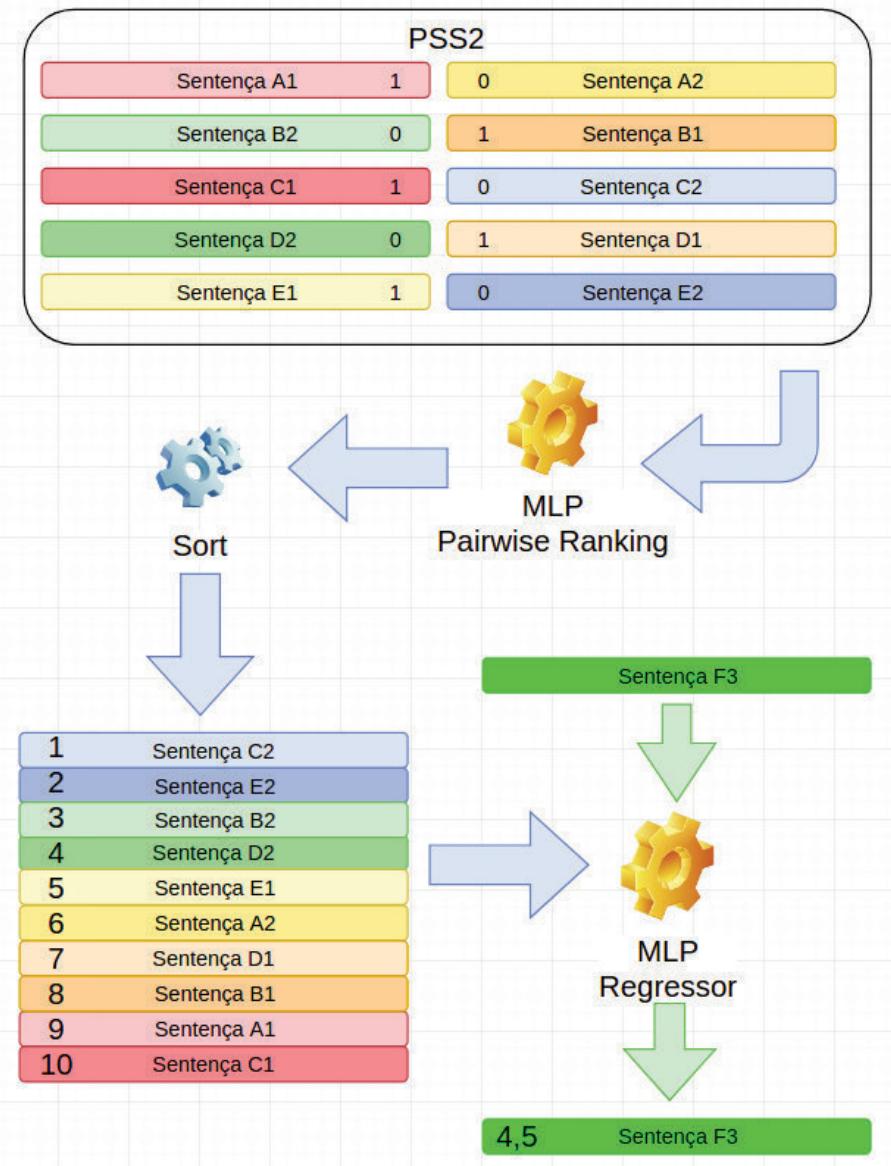


Tabela 3. Baselines e resultados do ranqueamento no PSS2

Modelo	Acurácia
Número de orações por sentença	41,28%
Média de sílabas por palavra de conteúdo	50,90%
Total de <i>tokens</i> por sentença	69,35%
SVMRank [Leal et al. 2018]	74,20%
MLP Pairwise Ranking	87,80%

giu 74,20% no mesmo córpus, conforme tabela 3, que também mostra outras 3 *baselines* simples nas primeiras linhas da tabela.

Fase 2 - Ranking global de sentenças

Uma vez obtido um modelo que conseguiu julgar razoavelmente bem a comple-

xidade das sentenças apresentadas em pares, ele foi utilizado para comparar todas as sentenças do PSS2, resultando em um *ranking* ordenado de 1 a 9.924. Algumas sentenças estão repetidas, quando aparecem em lados opostos do par, uma vez como a mais simples e outra vez como a mais complexa. Essas sentenças foram mantidas para uma validação adicional do resultado da ordenação, pois mesmo em lados diferentes, enquanto pares, no *ranking* global elas precisam estar próximas.

Conforme esperado, a maioria das sentenças do nível Original ficaram nas últimas posições do *ranking*, enquanto as primeiras posições foram preenchidas pelos níveis Natural e Forte. O primeiro terço do ranking ficou com 16% das sentenças originais, 30% das naturais e 55% das fortes. O último terço ficou com 52% de originais, 35% de naturais e 12% das fortes.

Fase 3 - Regressor

O *ranking* resultante da fase 2 foi normalizado entre 0 e 1 e utilizado para treinar uma segunda rede neural (MLP), também com 3 camadas. Porém, dessa vez, com apenas 189 neurônios na camada de entrada (apenas uma sentença) e predizendo a complexidade entre 0 e 1 no único neurônio de saída (utilizando ReLU), na forma de um regressor. O *dataset* foi dividido em 80% para treinamento e 20% para testes. Com todas as *features*, obtivemos uma raiz do erro quadrático médio (ou RMSE, em inglês) de 0,04 (MSE: 0,0017). Em seguida, foi aplicado o método de seleção de *features* **Permutation Importance** implementado no *eli5.sklearn*⁹ para escolher as 50 mais importantes¹⁰ e o regressor foi retreinado, obtendo uma RMSE de **0,06** (MSE: 0,0033). Foi implementada uma interface simples para validação no portal *open source* Simpligo (<https://simpligo.sidle.al>), na qual é possível entrar com um texto e conferir os valores preditos para cada sentença, numa escala de complexidade entre 0 e 100.

4. Teste de Robustez: generalização para outros gêneros de texto

O teste de robustez foi projetado para avaliar o desempenho do preditor de complexidade (em termos de *F-Measure*) em sentenças de outros gêneros diferentes do jornalístico no qual o modelo foi treinado. Foram escolhidas 500 sentenças do domínio rural conforme Tabela 4, de materiais selecionados com ajuda de uma pesquisadora da Embrapa Gado de Leite. Os materiais vieram dos gêneros instrucionais/procedimentais, técnicos e administrativos e foram agrupados para atender os quatro níveis de letramento do INAF 2018, sendo eles: rudimentar¹¹; elementar¹²; intermediário¹³ e proficiente¹⁴.

⁹https://eli5.readthedocs.io/en/latest/blackbox/permutation_importance.html

¹⁰A lista completa das *features* pode ser vista junto aos códigos fonte do trabalho em <https://github.com/sidleal/simpligo-ranker>

¹¹As cartilhas podem ser acessadas a partir dos seguintes links: <https://www.infoteca.cnptia.embrapa.br/handle/doc/1035506> e <https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1055203>.

¹²Os materiais de EaD vieram do espaço e-Campo (<https://www.embrapa.br/e-campo/>)

¹³Os comunicados técnicos (COT) podem ser acessados a partir dos links: <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1034878/1/COT77Teormatseca.pdf> e <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/594901/1/COT28Ensilagemdomilhoedosorgo.pdf>

¹⁴O Plano Diretor: <http://www.cnpms.embrapa.br/edital/PDU.pdf>.

As sentenças foram passadas pelo regressor e classificadas em 4 níveis, de acordo com a complexidade predita, sendo 1 o mais simples e 4 o mais complexo: Nível 1 (1-25), Nível 2 (26-50), Nível 3 (51-75) e Nível 4 (76-100). Posteriormente, elas foram avaliadas pela pesquisadora da Embrapa que anotou aquelas em que discordou do nível atribuído pelo regressor, atribuindo o nível que considerava apropriado. A Tabela 5 traz os resultados, sendo a *F-Measure* média obtida de **88,4%**.

Tabela 4. Sentenças selecionadas para teste de robustez do piloto

Publicação	Quantidade de Sentenças
Cartilhas de Ensilagem Milho e Sorgo	98
Curso de EaD de Silagem Capim	97
Curso de EaD de Silagens de milho e sorgo para produção de leite	95
Comunicado Técnico sobre Matéria Seca	61
Comunicado Técnico sobre Milho e Sorgo	91
Plano Diretor para Milho e Sorgo	58
Total	500

Quanto aos erros, 59,6% foram contíguos (nos quais o nível correto é imediatamente acima ou inferior ao predito) e 40,3% foram erros distantes. Os erros mais comuns aconteceram pela presença de termos que são simples no domínio rural, mas pouco frequentes nos demais domínios, e nas sentenças com pontuação diferente (por exemplo, as terminadas em dois pontos), o que vai exigir reavaliação das métricas utilizadas, incluindo novas métricas de natureza lexical para o domínio rural. Importante salientar que esses erros não são necessariamente do regressor, mas podem ter sido introduzidos por deficiências das ferramentas das etapas anteriores.

Tabela 5. Resultados do teste robustez

Nível	Precisão (%)	Recall (%)	<i>F-Measure</i> (%)
Nível 1	98,0	76,2	85,7
Nível 2	86,0	83,2	84,6
Nível 3	89,1	94,6	91,8
Nível 4	84,6	100,0	91,6
Média	89,4	88,5	88,4

5. Considerações Finais

Este trabalho apresentou uma evolução significativa para a abordagem da tarefa de Avaliação da Complexidade Sentencial para o português brasileiro, com um incremento de mais de 10% na acurácia sobre o melhor resultado anterior reportado em [Leal et al. 2018]. Também disponibilizou uma aplicação prática para o modelo, permitindo a avaliação das sentenças de um texto, além de provar sua capacidade de generalização para outros domínios. O teste de robustez demonstrou que o modelo desenvolvido pode ser útil no apoio da avaliação e simplificação dos materiais usados pela Embrapa, mesmo esses sendo de outros gêneros textuais. A análise de erros mostrou que novas métricas simples podem ajudar a aumentar o desempenho da tarefa. Como trabalhos futuros, pretendemos investigar a tarefa utilizando métodos de *Deep Learning* e *Transfer Learning*, além da inclusão de mais *features* no modelo, em especial as de rastreamento ocular reportadas no trabalho detentor do estado da arte da tarefa para a língua inglesa.

Referências

- Aluisio, S. and Gasperin, C. (2010). Fostering digital inclusion and accessibility: the Porsimples project for simplification of Portuguese texts. In *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas - Association for Computational Linguistics*, pages 46–53.
- Ambati, B. R., Reddy, S., and Steedman, M. (2016). Assessing relative sentence complexity using an incremental CCG parser. In *Proceedings of the North American Chapter of the Association for Computational Linguistics on Language technologies*, pages 1051–1057.
- Bick, E. (2000). *The Parsing System “Palavras”: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Aarhus University Press.
- Bosco, G. L., Pilato, G., and Schicchia, D. (2018). A neural network model for the evaluation of text complexity in Italian language: a representation point of view. In *Postproceedings of the 9th Annual International Conference on Biologically Inspired Cognitive Architectures (BICA 2018)*, pages 464–470.
- Brunato, D., Mattei, L. D., Dell’Orletta, F., Iavarone, B., and Venturi, G. (2018). Is this sentence difficult? Do you agree? In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2690–2699.
- Caseli, H. M., Pereira, T. F., Specia, L., Pardo, T. A. S., Gasperin, C., and Aluísio, S. M. (2009). Building a Brazilian Portuguese parallel corpus of original and simplified texts. *Advances in Computational Linguistics, Research in Computer Science (CICLing-2009)*, vol. 41:59–70.
- Cunha, A. L. V. (2015). *Coh-Metrix-Dementia: análise automática de distúrbios de linguagem nas demências utilizando Processamento de Línguas Naturais*. Master’s thesis, ICMC - USP, São Carlos - SP - Brasil.
- Dell’Orletta, F., Wieling, M., Cimino, A., Venturi, G., and Montemagni, S. (2014). Assessing the readability of sentences: Which corpora and features? *Proceedings of the Ninth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 163–173.
- Dell’Orletta, F., Montemagni, S., and Venturi, G. (2011). Read-it: Assessing readability of Italian texts with a view to text simplification. *Proceedings of the 2nd Workshop on Speech and Language Processing for Assistive Technologies*, pages 73–83.
- dos Santos, L. B., Duran, M. S., Hartmann, N. S., Candido, A., Paetzold, G. H., and Aluisio, S. M. (2017). A lightweight regression method to infer psycholinguistic properties for Brazilian Portuguese. In Ekštein, K. and Matoušek, V., editors, *Text, Speech, and Dialogue. TSD 2017. Lecture Notes in Computer Science*, volume 10415, pages 281–289. Springer, Cham.
- Dubay, W. H. (2007). *Smart Language: Readers, Readability, and the Grading of Text*. Impact Information, Costa Mesa, CA. ISBN: 1-4196-5439-X.
- Gonzalez-Garduño, A. V. and Søgaard, A. (2018). Learning to predict readability using eye-movement data from natives and learners. *Proceedings of the The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, pages 5118–5124.

- Howcroft, D. M. and Demberg, V. (2017). Psycholinguistic models of sentence processing improve sentence readability ranking. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 958–968.
- IPM (2018). *INAF Brasil 2018: Indicador de Alfabetismo Funcional - Resultados Preliminares*. Instituto Paulo Montenegro. Disponível em http://acaoeducativa.org.br/wp-content/uploads/2018/08/Inaf2018_Relat%C3%B3rio-Resultados-Preliminares_v08Ago2018.pdf.
- Leal, S. E., Duran, M. S., and Aluísio, S. M. (2018). A nontrivial sentence corpus for the task of sentence readability assessment in Portuguese. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 401–413.
- Magalhães, V. M. A., Bernardo, W. F., Diniz, F. H., dos Santos, K. C. L., Fonseca, L. M. G., Aluisio, S. M., and Leal, S. E. (2017). E-rural methodology: Contents elaborated according to the literacy level of the target audience. In *Twelfth Latin American Conference on Learning Technologies (LACLO)*, pages 1–9.
- Maziero, E. G., Pardo, T. A. S., and Aluísio, S. M. (2008). *Ferramenta de Análise Automática de Intelligibilidade de Córpus (AIC)*. NILC - ICMC-USP. Disponível em <http://www.nilc.icmc.usp.br/nilc/download/NILCTR0808-MazieroPardo.pdf>.
- Scarton, C. and Aluísio, S. (2010). Análise da inteligibilidade de textos via ferramentas de processamento de língua natural: adaptando as métricas do coh-metrix para o português. *Linguamática*, pages 45–62.
- Scarton, C., Paetzold, G. H., and Specia, L. (2018). Text simplification from professionally produced corpora. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 3504–3510.
- Singh, A. D., Mehta, P., Husain, S., and Rajkumar, R. (2016). Quantifying sentence complexity based on eye-tracking measures. *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity*, pages 202–212.
- Stajner, S., Ponzetto, S. P., and Stuckenschmidt, H. (2017). Automatic assessment of absolute sentence complexity. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, pages 4096–4102.
- Timm, L. B. (2018). *Looking at text simplification: Using eye tracking to evaluate the readability of automatically simplified sentences*. PhD thesis, Linköping University, Department of Computer and Information Science, Human-Centered systems, Linköping, Sweden.
- Vajjala, S. and Meurers, D. (2014). Assessing the relative reading level of sentence pairs for text simplification. *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 288–297.
- Vajjala, S. and Meurers, D. (2016). Readability-based sentence ranking for evaluating text simplification. *CoRR - Computer Research Repository*, Disponível em <http://arxiv.org/abs/1603.06009>.