

Proposta e Avaliação Linguística de Técnicas de Aumento de Dados

Arthur Scalercio¹, Cláudia Freitas²

¹Instituto de Computação – Universidade Federal Fluminense (UFF)

²Departamento de Letras – Pontifícia Universidade Católica (PUC-RJ)

arthurscalercio@id.uff.br, claudiafreitas@puc-rio.br

Abstract. *In Natural Language Processing (NLP), data augmentation consists of creating artificial training data for machine learning through textual transformations, aiming to improve the model's generalization capabilities and its performance in a range of downstream NLP tasks. Most studies on data augmentation techniques are evaluated based on the performance of the trained model with the artificially generated texts directly in the target task, without concern for linguistically evaluating the quality of the created texts. In this study, we propose two data augmentation techniques, evaluate the linguistic quality of the transformed texts, and demonstrate that the texts are well-constructed linguistically.*

Resumo. *Em Processamento de Linguagem Natural (PLN), aumento de dados consiste na criação artificial de dados para treinamento de modelos de aprendizado de máquina através de transformações nos textos, visando aumentar a capacidade de generalização de modelos e melhorar o desempenho em diversas tarefas de PLN. A maioria dos estudos sobre técnicas de aumento de dados são avaliados pelo desempenho do modelo treinado com os textos artificiais diretamente na tarefa alvo, não se preocupando em avaliar linguisticamente a qualidade dos textos criados. Nesse estudo, propomos duas técnicas de aumento de dados, avaliamos a qualidade linguística dos textos transformados e mostramos que os textos são linguisticamente bem construídos.*

1. Introdução

Nem todo problema de aprendizado é solucionado com aumento de dados (AD). Porém, a quantidade de dados para treinamento tem sido fundamental para o treinamento com sucesso de grandes modelos de linguagem [Radford et al. 2019, Devlin et al. 2019, Brown et al. 2020, Souza et al. 2020]. Há diversas situações em que se mostra vantajoso o uso de dados artificialmente construídos. Por exemplo, [Krishna et al. 2020] e [Lai et al. 2021] mostram que a adoção de paráfrases artificialmente contruídas melhora o desempenho de modelos neurais em diversas tarefas de transferência de estilo, como imitação autoral [Xu et al. 2012] e transferência de sentimento [Li et al. 2018]. A primeira consiste em converter o estilo de uma sentença para o estilo de determinado autor. A segunda consiste em converter a sentença para um sentimento diferente, mas mantendo seu assunto principal.

Normalmente, o sucesso da estratégia de aumento é medido avaliando-se o desempenho do modelo treinado nos dados aumentados em comparação com o modelo treinado nos dados não-aumentados, sem se preocupar se os textos gerados são bem formados. Diferentemente, esse estudo realiza uma avaliação intrínseca dos dados gerados, não avaliando se o aumento beneficia o desempenho de modelos em outras tarefas de PLN.

De forma exploratória, propõem-se, implementam-se e avaliam-se duas transformações para AD. Busca-se manter, linguisticamente, a correção e a naturalidade das frases criadas. Usando a gramática *Universal Dependency* (UD) [de Marneffe et al. 2021], movem-se sintagmas adverbiais que aparecem antes do respectivo verbo que modificam para depois dos blocos de sujeito, verbo e objeto. Usando o mesmo raciocínio, deslocam-se orações subordinadas adverbiais que aparecem antes do respectivo verbo que modificam para o final da frase, após sujeito, verbo e objetos. Três livros da coleção *Literatura para Todos* (CLPT) foram usados para avaliar os métodos. Excetuando quando o anotador nos fornece uma árvore sintática errada, as frases geradas mostraram-se bem construídas.

2. Método

O advérbio é fundamentalmente um modificador do verbo, podendo também reforçar o sentido de um adjetivo ou de um outro advérbio. Além disso, salienta-se que alguns advérbios aparecem, não raro, modificando toda a oração [Velo 2013, Costa 2008, Bechara 2001]. Neste último emprego, vêm geralmente destacados no início da oração, de cujos termos se separam por uma pausa nítida, marcada na escrita por vírgula. Quando movidos de lugar, geralmente podem ocupar qualquer outra posição na sentença, e até mesmo separar sujeito-verbo e verbo-objeto. Por outro lado, modificadores adverbiais internos à sentença geralmente não são marcados por pontuação e possuem uma posição mais fixa em relação ao termo que modificam.

Nossa estratégia de AD consiste em mudar de lugar estruturas adverbiais modificadoras cuja posição na sentença seja flexível. Para tanto, duas estruturas com função de advérbio foram escolhidas. Selecionaram-se as frases em que há uma oração subordinada adverbial aparecendo antes do respectivo verbo modificado por ela e as frases em que há um sintagma adverbial aparecendo antes do respectivo verbo modificado por ele. Em ambas as transformações, gera-se uma paráfrase usando as mesmas palavras da sentença de origem, mas movendo a estrutura linguística alvo para depois de sujeito, verbo e objetos.

Para alcançar esse objetivo, anotamos os textos conforme a gramática UD, usando o anotador sintático UDPipe [Straka et al. 2016], que fornece árvores de dependência que nos permitem encontrar as duas estruturas. Os rótulos sintáticos OBL e ADVCL da UD nos trazem, respectivamente, núcleos de sintagmas adverbiais e de orações subordinadas adverbiais. Uma vez encontrados, move-se a palavra correspondente ao núcleo da estrutura juntamente com todos os seus filhos na árvore de dependência para frente da frase, após os termos que constituem o sujeito, o verbo e os objetos. Mais detalhes do algoritmo que faz a translação encontram-se no Apêndice A.1. A árvore de dependência da sentença criada é igual à da sentença de entrada, ou seja, o método proposto não muda nem as palavras existentes na sentença, nem a função sintática exercida por cada uma delas. A ideia da translação do sintagma adverbial está ilustrada na Figura 1. Por último, um pós-processamento dos textos, descrito no Apêndice A.2, é feito para ajustar pontuações.

Em o começo , Cláudia até achou delicioso ; em a segunda refeição de bolo de laranja , já estava enjoada .

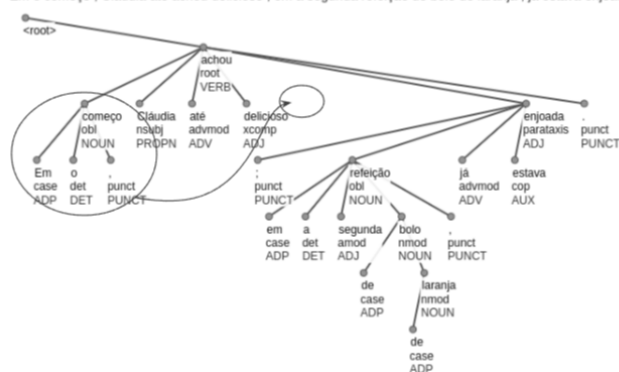


Figura 1. Translação do sintagma adverbial

3. Experimentos e Resultados

Para avaliar nosso método de AD, selecionaram-se três livros da coleção Literatura para Todos (CLPT): *Leão, o pardo*; *Madalena*; e *Tubarão com faca nas costas*. A escolha foi motivada pelo fato de eles serem obras literárias publicamente disponíveis¹ e destinados teoricamente a pessoas com habilidades de leitura iniciais. As estatísticas do conjunto de dados e a quantidade de análises qualitativas realizadas constam na Tabela 1.

Tabela 1. Estatísticas do conjunto de dados e dos experimentos

Rótulo	Livro	#Frases	#tokens	Anotações Avaliadas	Frases Avaliadas (1)	Frases Transformadas	Transformadas avaliadas (2)	Interseção de (1) e (2)
ADVCL	Leopardo	997	25027	20	20	88	20	18
	Madalena	1413	19655	20	20	92	20	20
	Tubarão	650	11678	20	20	99	20	20
	Total	3060	56360	60	60	279	60	58
OBL	Leopardo	997	25027	20	16	220	20	12
	Madalena	1413	19655	20	13	125	20	11
	Tubarão	650	11678	20	15	121	20	13
	Total	3060	56360	60	44	466	60	36

3.1. Análise de erros da anotação automática

Para cada uma das estruturas linguísticas procuradas, calculou-se a precisão do anotador UDPipe. Este anotador morfossintático foi escolhido pela facilidade de uso. Para ambas as estruturas linguísticas analisadas, 60 amostras foram aleatoriamente escolhidas, sendo 20 de cada livro. Para o rótulo ADVCL a precisão foi de 90%, enquanto para o OBL foi de 83.3%. Considerou-se que o anotador acerta se, ao mesmo tempo, classifica corretamente o rótulo sintático UD da palavra (ADVCL ou OBL, conforme o caso) e escolhe corretamente o pai da palavra na árvore sintática. Tabela 2 mostra os erros cometidos pelo anotador conforme essas dimensões. Mais detalhes sobre essa análise no Apêndice A.4.

Tabela 2. Análise da anotação sintática automática

Livro	Orações adverbiais antes do verbo			Sintagmas adverbiais antes do verbo		
	Erro Classificação	Erro Referência	Erro Ref. e Classif.	Erro Classificação	Erro Referência	Erro Ref. e Classif.
Leo-o-pardo	1	1	0	2	0	0
Madalena	1	0	0	1	0	3
Tubarão	1	0	2	2	1	1
Total	3	1	2	5	1	4

¹ <http://www.dominiopublico.gov.br/>

3.2. Análise da qualidade dos dados aumentados

Avaliaram-se qualitativamente 60 frases para cada estratégia de AD, sendo 20 de cada livro. Apesar de a maioria das frases analisadas terem como origem as mesmas cujas anotações foram avaliadas na Subseção 3.1, alguns textos aumentados analisados vieram de frases que não tiveram a anotação avaliada anteriormente, devido a alguns filtros inseridos na seleção das sentenças a serem aumentadas. Detalhes do processo de seleção das frases analisadas encontram-se no Apêndice A.3. Os resultados constam na Tabela 3.

Tabela 3. Análise dos dados aumentados

Livro	Orações adverbiais antes do verbo		Sintagmas adverbiais antes do verbo	
	#Frases Mal-formadas	#Análises	#Frases Mal-formadas	#Análises
Leo-o-pardo	2	20	4	20
Madalena	1	20	6	20
Tubarão	3	20	5	20
Total	6	60	15	60

Dentre as frases analisadas, todas as mal-formadas foram causadas por erros na anotação sintática da frase original. Uma análise das frases mal-formadas encontra-se no Apêndice A.5. Amostras das sentenças bem formadas e analisadas constam no Apêndice A.6. O código que permite a reproducibilidade integral dos experimentos encontra-se disponível².

4. Trabalhos Relacionados

[Bayer et al. 2022] realizou uma pesquisa sobre métodos de AD voltados para a tarefa de classificação e mostrou técnicas que atuam no nível do caractere, da palavra, da frase e do documento, e podem ser baseadas em regras ou em algoritmos de aprendizado de máquina. Dentre as baseadas em regras que criam dados com transformações sintáticas, [Min et al. 2020] melhora o resultado de modelos na tarefa de inferência textual. Para o português brasileiro, [Menezes et al. 2023] mostra que modelos para simplificação textual treinados com dados aumentados apresentam maior generalização quando aplicados a outros domínios. [Şahin and Steedman 2018] cria dados, através da seleção de pedaços de frases e da inversão da ordem de aparecimento de alguns termos sintáticos nas frases. Esses dados artificiais beneficiaram o desempenho de um modelo neural na tarefa de anotação de desambiguação morfossintática (em inglês, part-of-speech (POS) tagging), em diversas línguas com poucos recursos. De forma parecida, nossa abordagem cria dados através de transformações. No entanto, estas são mais complexas e levam em conta as peculiaridades do posicionamento de modificadores adverbiais na língua portuguesa.

5. Conclusão e Trabalhos Futuros

O presente estudo propôs e avaliou linguisticamente duas estratégias para a geração de paráfrases. Embora a amostra analisada seja pequena, a ideia se mostrou promissora e é possível estendê-la para mover outras estruturas linguísticas que possam estar fora da ordem canônica do português, como sujeito e objeto. Vimos que o anotador UDPipe, para sentenças com poucos verbos e sem discurso direto, possui uma boa qualidade. Planeja-se confirmar a efetividade do método avaliando se o aumento beneficia modelos de aprendizado em outras tarefas, como na de simplificação textual, onde frases na ordem canônica e fora da ordem canônica podem ser usadas em um treinamento supervisionado.

²<https://github.com/scalercio/aumento-dados-ud>

Referências

- Bayer, M., Kaufhold, M.-A., and Reuter, C. (2022). A survey on data augmentation for text classification. *ACM Computing Surveys*, 55(7):1–39.
- Bechara, E. (2001). *Moderna Gramática Portuguesa*. Lucerna.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.
- Costa, J. (2008). *O Advérbio em Português Europeu*. Colibri, Lisboa.
- de Marneffe, M.-C., Manning, C. D., Nivre, J., and Zeman, D. (2021). Universal Dependencies. *Computational Linguistics*, 47(2):255–308.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Krishna, K., Wieting, J., and Iyyer, M. (2020). Reformulating unsupervised style transfer as paraphrase generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 737–762, Online. Association for Computational Linguistics.
- Lai, H., Toral, A., and Nissim, M. (2021). Generic resources are what you need: Style transfer tasks without task-specific parallel training data. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4241–4254, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Li, J., Jia, R., He, H., and Liang, P. (2018). Delete, retrieve, generate: a simple approach to sentiment and style transfer. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, Volume 1 (Long Papers)*, pages 1865–1874. Association for Computational Linguistics.
- Menezes, L., Paes, A., and Finatto, M. (2023). Abordagem baseada em aumento de dados para avaliação automática de leitura. *Domínios de Linguagem*, 17:e1721.
- Min, J., McCoy, R. T., Das, D., Pitler, E., and Linzen, T. (2020). Syntactic data augmentation increases robustness to inference heuristics. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2339–2352, Online. Association for Computational Linguistics.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language models are unsupervised multitask learners.
- Şahin, G. G. and Steedman, M. (2018). Data augmentation via dependency tree morphing for low-resource languages. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5004–5009, Brussels, Belgium. Association for Computational Linguistics.

- Souza, F., Nogueira, R., and Lotufo, R. (2020). Bertimbau: Pretrained bert models for brazilian portuguese. In Cerri, R. and Prati, R. C., editors, *Intelligent Systems*, pages 403–417, Cham. Springer International Publishing.
- Straka, M., Hajič, J., and Straková, J. (2016). UDPipe: Trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4290–4297, Portorož, Slovenia. European Language Resources Association (ELRA).
- Veloso, R. (2013). *Gramática do Português*, volume 2, chapter Advérbio e Sintagma Adverbial, pages 1569–1684. Fundação Calouste Gulbenkian.
- Xu, W., Ritter, A., Dolan, B., Grishman, R., and Cherry, C. (2012). Paraphrasing for style. In Kay, M. and Boitet, C., editors, *COLING 2012, 24th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, 8-15 December 2012, Mumbai, India*, pages 2899–2914. Indian Institute of Technology Bombay.

A. Apêndice

A.1. Detalhes do Algoritmo de Translação de Estruturas Linguísticas

Para encontrar as frases passíveis de serem transformadas pela nossa técnica, busca-se a palavra que esteja rotulada em UD como OBL (núcleo do sintagma adverbial) ou ADVCL (núcleo da oração subordinada adverbial) e cujo pai dela na árvore de dependência sintática esteja localizado a frente dela. Caso a frase não tenha essa estrutura, nada é feito com ela. Caso alguma dessas estruturas esteja presente, a técnica de aumento será aplicada nela. O núcleo da estrutura juntamente com seus filhos na árvore serão levados para uma posição após o verbo e objetos, caso hajam. Para o sucesso linguístico da sentença gerada, é necessário que a estrutura movida seja inserida não só após as palavras que compõem sujeito, verbo e objetos, mas também após alguns rótulos que indicam palavras que estão vinculadas diretamente ao verbo, tais como verbos de ligação, verbos auxiliares, conjunções, pronomes reflexivos e pontuações. Para englobar tais estruturas, seguem os rótulos em UD, vinculados ao pai da estrutura a ser movida, que devem ser considerados: nsubj, advmod, iobj, obj, obl, advcl, aux, punct, cop, mark, conj, csubj, xcomp, expl. Assim, a estrutura movida é inserida após a ocorrência do último de qualquer desses rótulos.

Para preservar a qualidade dos textos gerados, o algoritmo também não permite que os advérbios modificadores sejam movidos de posições anteriores para posteriores em relação a algumas pontuações (".", ":", "!", "?"). Além disso, durante o processo de translação dos blocos de sintagmas adverbiais e de orações adverbiais, caso haja outro bloco com a mesma função sintática do bloco movido e que tenha índice maior que seu pai na árvore, o bloco movido será inserido antes desse bloco.

Pontua-se que quando o termo a ser movido inicia a frase original, insere-se na sentença criada uma vírgula imediatamente antes da posição onde o termo é inserido. Por fim, quando se busca os sintagmas adverbiais para transformação, não se consideram as palavras anotadas como OBL e que estejam anotadas como pronome relativo (ex: "que"). Trata-se de ajustes que tiveram que ser feitos no código para permitir uma geração de textos adequada. Por não ser alvo do presente estudo, o algoritmo não se preocupou em converter letras maiúsculas para minúsculas e vice-versa.

A.2. Pós-processamento

Após o algoritmo principal, um simples pós-processamento das frases geradas é realizado para ajustar pontuações que passaram a aparecer em sequência. Seguem os ajustes:

- DE , , → PARA ,
- DE , . → PARA .
- DE , ; → PARA ;
- DE , ? → PARA ?
- DE , ! → PARA !

A.3. Detalhes do processo de seleção das frases analisadas

Para a realização da análise de erros do anotador UDPipe, escolheram-se aleatoriamente sentenças dos três livros da CLPT, e nessas sentenças buscaram-se as etiquetas OBL ou ADVCL (de acordo com a estrutura linguística analisada) que apareciam antes do termo

que modificavam. Pontua-se que havia frases que possuíam mais de uma etiqueta OBL na situação procurada. Essa é a razão que fez com que essa análise utilizasse menos de 20 (vinte) frases por livro, apesar de terem sido analisadas 20 estruturas linguísticas por livro (ver Tabela 1).

Para a escolha das frases aumentadas a serem avaliadas linguisticamente, pegaram-se as sentenças geradas a partir das frases que tiveram suas anotações avaliadas. Porém, como o processo de avaliação da anotação usou menos de 20 frases por livro, escolheram-se aleatoriamente as restantes até completar 20 (vinte) frases por livro.

A.4. Detalhes dos erros do anotador automático

Os erros de classificação e de referência constantes na Tabela 2 encontram-se mais detalhados nas Tabelas 4 e Tabelas 5. Percebe-se uma dificuldade do anotador em lidar com discurso direto, o qual deveria ser anotado como PARATAXIS, pois não há uma conexão formal, explícita, entre as duas orações.

Tabela 4. Erros do Anotador - ADVCL

Tipo de Erro	Rótulo Correto	#Qtd	%
Classificação	PARATAXIS	1	16.6%
	ADVMOD	1	16.6%
	XCOMP	1	16.6%
Classif. e	ACL:REL	1	16.6%
Referência	PARATAXIS	1	16.6%
Referência		1	16.6%
Total		6	100%

Tabela 5. Erros do Anotador - OBL

Tipo de Erro	Rótulo Correto	#Qtd	%
Classificação	IOBJ	3	30%
	PARATAXIS	1	10%
	MARK	1	10%
Classif. e	ROOT	3	30%
Referência	PARATAXIS	1	10%
Referência		1	10%
Total		10	100%

A.5. Análise das frases aumentadas mal-formadas

Tabela 6, Tabela 7 e Tabela 8 listam algumas frases mal-formadas usando nosso método para os livros “*Léo, o pardo*”, “*Madalena*” e “*Tubarão com faca nas costas*”, respectivamente. Da análise, nota-se, de um modo geral, que estas sentenças aumentadas são provenientes de frases com discurso direto ou com diversas orações.

A.6. Amostras de frases originais e aumentadas bem-formadas

Tabela 9 e Tabela 10 listam algumas frases aumentadas do livro “*Léo, o pardo*” e consideradas bem-formadas. Tabela 11 e Tabela 12 listam algumas frases bem-formadas do livro “*Madalena*”. E Tabela 13 e Tabela 14 listam algumas frases bem-formadas do livro “*Tubarão com faca nas costas*”.

Tabela 6. Frases mal-formadas do livro “Léo, o pardo”

TIPO	Translação Oração adverbial
Original	“ Mãe , a professora disse que toda mãe deixa de comer pra dar pro filho , se for preciso , igual alguns bichos ” .
Aumentada	“ Mãe , a professora disse que toda mãe deixa de comer pra dar pro filho , igual alguns bichos se for preciso ” .
Original	“ Como , se eu não fiz colégio bom e não sei química e nem física ? ” Ele jogou a deixa , primeiro igual o professor Lázaro .
Aumentada	“ Como , jogou se eu não fiz colégio bom e não sei química e nem ? ” Ele física a deixa , primeiro igual o professor Lázaro .
TIPO	Translação Sintagma adverbial
Original	Pra gente ela sempre aconselhava , “ não chora tanto , menino , que lágrima é uma coisa que seca , quando a alma mais precisar de ela , aí não desce mais ” .
Aumentada	ela sempre aconselhava , Pra gente , “ não chora tanto , menino , que lágrima é uma coisa que seca , quando a alma mais precisar de ela , aí não desce mais ” .
Original	Em a primeira calça ela foi meio sem jeito , mediu a parte de a frente , fez a parte de trás maior por causa de a bunda , emendou as duas , o cliente gostou e ela passou a fazer roupa pra fora .
Aumentada	ela foi meio , Em a primeira calça sem jeito , mediu a parte de a frente , fez a parte de trás maior por causa de a bunda , emendou as duas , o cliente gostou e ela passou a fazer roupa pra fora .

Tabela 7. Frases mal-formadas do livro “Madalena”

TIPO	Translação Oração adverbial
Original	Porque pensar ela podia , e pensando ela podia tudo .
Aumentada	podia , Porque pensar ela , e pensando ela podia tudo .
TIPO	Translação Sintagma adverbial
Original	- Mais ou menos , mainha - e ele olhou para Dulce .
Aumentada	- mainha Mais ou menos , - e ele olhou para Dulce .
Original	- É muito linda , Dinda ... - e , com a melhor cara de cachorro de porta de açougue que sabia fazer - Um dia me dá ela ?
Aumentada	- É muito linda , Dinda ... - e - Um dia me dá ela , com a melhor cara de cachorro de porta de açougue que sabia fazer ?
Original	Quando amanheceu , viu Francisco dormindo em o banco de trás de o carro de dona Madalena , o mesmo banco onde tantas e tantas vezes ... Bateu em o vidro .
Aumentada	Quando amanheceu , viu Francisco dormindo em o banco de trás de o carro de dona Madalena , o mesmo banco onde Bateu tantas e tantas ... vezes em o vidro .
Original	E nem por ela ele fez alguma coisa , quando Nelson a entregou a Manuel em troca de uma braçada de terras .
Aumentada	E ele fez alguma coisa , quando Nelson a entregou a Manuel em troca de uma braçada de terras nem por ela .

Tabela 8. Frases mal-formadas do livro “Tubarão com faca nas costas”

TIPO	Translação Oração adverbial
Original	E assim fui , feliz , por um pedaço de o percurso , com aquela criança pequena me confortando mais de o que eu a ela , até que - surpresa - a mãe decidiu descer bem antes de onde disse que desceria .
Aumentada	E assim fui , feliz , por um pedaço de o percurso , com aquela criança pequena me confortando mais de o que eu a ela , até que a mãe decidiu descer bem antes de onde disse que desceria - surpresa - .
Original	Só deu tempo de ele parar o ônibus . “Não cheguei a conhecer os pais de Santiago , mas toda vez que lhe faço uma visita , seja para jogarmos xadrez ou para discutirmos um novo texto , ele dá um jeito de falar em ambos , principalmente em a mãe , que era com quem ele mais se identificava .
Aumentada	Só deu tempo de ele parar o ônibus . “Não cheguei a conhecer os pais de Santiago , ele dá um jeito de falar em ambos , principalmente em a mãe , que era com quem ele mais se identificava , mas toda vez que lhe faço uma visita , seja para jogarmos xadrez ou para discutirmos um novo texto .
TIPO	Translação Sintagma adverbial
Original	É graças a esse bichano que tenho para mim que nenhum animal de estimação é demasiadamente esperto , senão não aceitariam viver em uma família como a nossa .
Aumentada	É que tenho graças a esse bichano para mim que nenhum animal de estimação é demasiadamente esperto , senão não aceitariam viver em uma família como a nossa .
Original	Em suma : o amor deve estar incrustado em tudo e em todos .
Aumentada	deve , Em : o amor suma estar incrustado em tudo e em todos .

Tabela 9. Frases bem-formadas do livro “Léo, o pardo”

TIPO	Translação oração adverbial
Original	Revoltadas com a condenação , falaram pra juíza : “ então manda matar de uma vez ” .
Aumentada	falaram pra juíza , Revoltadas com a condenação : “ então manda matar de uma vez ” .
Original	”Pra fazer cinema , primeiro se aprende fotografia , literatura , se vê muito filme , música .
Aumentada	”primeiro se aprende fotografia , literatura Pra fazer cinema , se vê muito filme , música .
Original	Voltando de um trabalho como monitor em a biblioteca de o Instituto de Matemática , reencontrei a Sabrina , professora de o curso de cinema de Santo André , e recebi de ela os parabéns por ter passado em a USP .
Aumentada	reencontrei a Sabrina , professora de o curso de cinema de Santo André , Voltando de um trabalho como monitor em a biblioteca de o Instituto de Matemática , e recebi de ela os parabéns por ter passado em a USP .
Original	A máquina de lavar roupas quebrou em a única vez que eu usei ; ”se quebrar de novo será descontado de o seu salário ” .
Aumentada	A máquina de lavar roupas quebrou em a única vez que eu usei ; “ será descontado de o seu salário se quebrar de novo ” .
Original	Sentado em o próximo 109 ônibus , fiquei pensando o que aconteceria se não eu fosse de a USP e ainda estivesse desempregado .
Aumentada	fiquei pensando o que aconteceria se não eu fosse de a USP e ainda estivesse desempregado , Sentado em o próximo 109 ônibus .
Original	Em o trabalho , a Dona Nega gritava de lá : “ alguém quer bala ? ” E a Maria de o Antônio Joana , depois de aceitar , gritava em o outro canto , cuspindo : “ cruz credo , nunca comi bala de espinho ! ” E a Dona Nega caía em a risada : “ é de coco , criatura ” .
Aumentada	Em o trabalho , a Dona Nega gritava de lá : “ alguém quer bala ? ” E a Maria de o Antônio Joana gritava em o outro canto , depois de aceitar , cuspindo : “ cruz credo , nunca comi bala de espinho ! ” E a Dona Nega caía em a risada : “ é de coco , criatura ” .

Tabela 10. Frases bem-formadas do livro “Léo, o pardo”

TIPO	Translação Sintagma adverbial
Original	Anos mais tarde , com esse tal Cambraia já morto , o filho de ele mais moço , o Geirson Cambraia , descobriu um monte de papel amarelado em uma gaveta e quase morreu de desgosto ; eram os documentos de tudo quanto era terra alheia que se passava por terra de os Cambraia .
Aumentada	com esse tal Cambraia já morto , o filho de ele mais moço , o Geirson Cambraia , descobriu um monte de papel amarelado em uma gaveta , Anos mais tarde , e quase morreu de desgosto ; eram os documentos de tudo quanto era terra alheia que se passava por terra de os Cambraia .
Original	Enquanto isso , em a nossa curiosidade que nunca morria , ela ia contando detalhes de o tempo antigo .
Aumentada	em a nossa curiosidade que nunca morria , ela ia contando detalhes de o tempo antigo , Enquanto isso.
Original	Em a USP , a maioria de os homens e mulheres negras , e são muitos , vestem roupas vermelhas , verdes , alaranjadas , azuis ou pretas e cortam gramas , capinam jardins , guardam portas e servem comidas em bandejas pros alunos brancos , em o restaurante central , ou laboram em eternas construções e depois vão pra casa , ali de o lado .
Aumentada	a maioria de os homens e mulheres negras , e são muitos , vestem roupas vermelhas , verdes , alaranjadas , azuis ou pretas , Em a USP , e cortam gramas , capinam jardins , guardam portas e servem comidas em bandejas pros alunos brancos , em o restaurante central , ou laboram em eternas construções e depois vão pra casa , ali de o lado .
Original	Com o cargo ganharia mil e seiscentos reais e faria faculdade de cinema .
Aumentada	ganharia mil e seiscentos reais , Com o cargo e faria faculdade de cinema .

Tabela 11. Frases bem-formadas do livro “Madalena”

TIPO	Translação oração adverbial
Original	Enquanto Madalena se vestia , Rubina puxou sua maleta de couro de baixo de a cama e foi enfiando em ela o conteúdo de a trouxa : uma meia com algum dinheiro , dois vestidos , uma camisola , um casaco , toalhinhas , cueiros , dois conjuntos de Chiquinho .
Aumentada	Rubina puxou sua maleta de couro de baixo de a cama , Enquanto Madalena se vestia , e foi enfiando em ela o conteúdo de a trouxa : uma meia com algum dinheiro , dois vestidos , uma camisola , um casaco , toalhinhas , cueiros , dois conjuntos de Chiquinho .
Original	E , apesar de nunca ter destrutado Francisco , Maria tampouco gostava de esse namoro ; sempre que podia , trazia a a tona esse assunto : ”Cuidado com esse Francisco .
Aumentada	E Maria tampouco gostava de esse namoro , apesar de nunca ter destrutado Francisco ; sempre que podia , trazia a a tona esse assunto : “ Cuidado com esse Francisco .
Original	Além de ser a encarregada de todos os não disciplinadores , Dulce era , a os olhos de Cláudia , infinitamente menos interessante que Madalena .
Aumentada	Dulce era , a os olhos de Cláudia , infinitamente menos interessante que Madalena , Além de ser a encarregada de todos os não disciplinadores .
Original	Se estava se separando por a segunda vez era por uma fatalidade .
Aumentada	era por uma fatalidade , Se estava se separando por a segunda vez .
Original	Não perdia tempo discutindo , mas tinha lá suas idéias , e quando cismava com uma coisa , não sossegava até conseguir o que queria .
Aumentada	Não perdia tempo discutindo , mas tinha lá suas idéias , e não sossegava quando cismava com uma coisa , até conseguir o que queria .

Tabela 12. Frases bem-formadas do livro “Madalena”

TIPO	Translação Sintagma adverbial
Original	Em a época , Dulce ainda pesava menos de quarenta quilos por a depressão que havia sofrido com a morte repentina de o pai , ocorrida dois anos antes de o namoro começar .
Aumentada	Dulce ainda pesava menos de quarenta quilos , Em a época , por a depressão que havia sofrido com a morte repentina de o pai , ocorrida dois anos antes de o namoro começar .
Original	Em o final de cada mês , eles deixavam todo o salário em a venda de a própria fazenda , e em o dia seguinte a o pagamento , fatalmente , já começavam outra dívida .
Aumentada	eles deixavam todo o salário , Em o final de cada mês , em a venda de a própria fazenda , e em o dia seguinte a o pagamento , fatalmente , já começavam outra dívida .
Original	Aquela noite em que Francisco foi arrancado de a cama que dividia com a mãe por um sujeito imenso - ou , por o menos , imenso para os quatro anos que ele tinha - que caiu em cima de ela .
Aumentada	Aquela noite em que Francisco foi arrancado de a cama que dividia com a mãe por um sujeito imenso - ou imenso , por o menos , para os quatro anos que ele tinha - que caiu em cima de ela .
Original	Porque , em casa , a avó não dava trela para ninguém - saía , trabalhava , voltava , viajava , saía de novo .
Aumentada	Porque a avó não dava trela , em casa , para ninguém - saía , trabalhava , voltava , viajava , saía de novo .
Original	- Em a horta - respondeu , recolhendo a garrafa de cachaça e os copos ; verificou se o vidro de o tampo de a mesinha não estava trincado .
Aumentada	respondeu , recolhendo a garrafa de cachaça e os copos ; - Em a horta - verificou se o vidro de o tampo de a mesinha não estava trincado .

Tabela 13. Frases bem-formadas do livro “Tubarão com faca nas costas”

TIPO	Translação oração adverbial
Original	Quando o ônibus chegou , ela pediu (ainda ignorando Maira e demais pessoas) que a ajudasse com o pequeno .
Aumentada	ela pediu , Quando o ônibus chegou , (ainda ignorando Maira e demais pessoas) que a ajudasse com o pequeno .
Original	Antes de perdê-lo de vista , vejo o inteiramente só , em o meio de a rua .
Aumentada	vejo o inteiramente só , em o meio de a rua , Antes de perdê-lo de vista .
Original	Passadas duas ou três cervejas e umas boas risadas , começamos a imaginar o que teria acontecido a a moça , dona de o decote .
Aumentada	começamos a imaginar o que teria acontecido a a moça , dona de o decote , Passadas duas ou três cervejas e umas boas risadas .
Original	Quem o encontrou foi meu pai , que , após comerem , tinha saído para caminhar e conhecer a propriedade que o amigo comprara a fim de dar um pouco de paz a a família .
Aumentada	Quem o encontrou foi meu pai , que tinha saído , após comerem , para caminhar e conhecer a propriedade que o amigo comprara a fim de dar um pouco de paz a a família .
Original	Aproveitando se de esse descuido , um pequeno tumor começou a trabalhar em o seu ardiloso intento .
Aumentada	um pequeno tumor começou a trabalhar em o seu ardiloso intento , Aproveitando se de esse descuido .

Tabela 14. Frases bem-formadas do livro “Tubarão com faca nas costas”

TIPO	Translação Sintagma adverbial
Original	Para tu teres uma idéia , em a mesma época em que o Mariozinho morreu , o Élcio , um grande camarada que tu não chegaste a conhecer , foi morar em Florianópolis com a mulher .
Aumentada	Para tu teres uma idéia , o Élcio , um grande camarada que tu não chegaste a conhecer , foi morar em a mesma época em que o Mariozinho morreu , em Florianópolis com a mulher .
Original	Para sorte não de o gato , morávamos em uma casa de dois andares .
Aumentada	morávamos , Para sorte não de o gato , em uma casa de dois andares .
Original	E , com uma imagem clara de a minha mãe em a saudade , contei uma história de como (quase) me perdi .
Aumentada	E contei uma história de como (quase) me perdi , com uma imagem clara de a minha mãe em a saudade .
Original	Em seguida , tomo um café , volto para o quarto e leio o que me der vontade .
Aumentada	tomo um café , Em seguida , volto para o quarto e leio o que me der vontade .
Original	Dia de esses , quando fomos a Rio Grande , cidade onde nasceste , enquanto me esperavas engraxar os sapatos para a formatura a a qual havíamos sido convidados , o engraxate , um senhor que regulava em idade contigo , em os colocou a par de o que acontecera a um semnúmero de pessoas , todas conhecidas tuas .
Aumentada	quando fomos a Rio Grande , cidade onde nasceste , enquanto me esperavas engraxar os sapatos para a formatura a a qual havíamos sido convidados , o engraxate , um senhor que regulava em idade contigo , em os colocou , Dia de esses , a par de o que acontecera a um semnúmero de pessoas , todas conhecidas tuas .