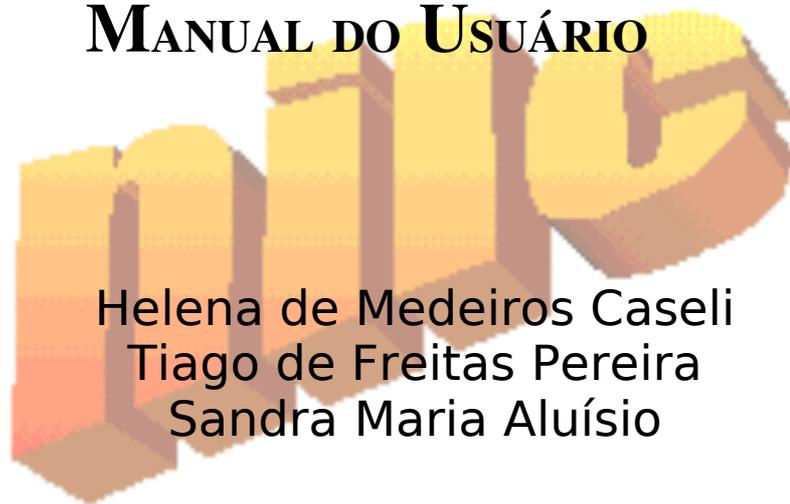


Universidade de São Paulo - USP
Universidade Federal de São Carlos - UFSCar
Universidade Estadual Paulista - UNESP

EDITOR DE ANOTAÇÃO DE SIMPLIFICAÇÃO: MANUAL DO USUÁRIO



Helena de Medeiros Caseli
Tiago de Freitas Pereira
Sandra Maria Aluísio

NILC-TR-08-10

Julho, 2008

Série de Relatórios do Núcleo Interinstitucional de Linguística
Computacional
NILC - ICMC-USP, Caixa Postal 668, 13560-970 São Carlos, SP,
Brasil

Editor de Anotação de Simplificação: Manual do Usuário

Índice

1.Introdução.....	1
2.Criação de uma nova produção.....	3
3.Edição de uma produção.....	4
4.Simplificação Natural.....	5
Operações de Simplificação.....	7
O Léxico.....	10
Finalizando uma Simplificação Natural.....	13
5.Simplificação Forte.....	13
O Sintático.....	14
Finalizando uma Simplificação Forte.....	16
6.Referências Bibliográficas.....	17

1. Introdução

Esse documento visa apresentar, de maneira sucinta, as funcionalidades básicas do *Editor de Anotação de Simplificação* textual desenvolvido no escopo do projeto PorSimples e disponível em <http://caravelas.icmc.usp.br/annotador/index>.

A simplificação textual, objeto de estudo no projeto PorSimples, está pensada para ser realizada em duas etapas. Primeiramente, o(a) anotador(a) terá a liberdade de realizar uma simplificação mais “natural”, sem regras muito rígidas a serem seguidas. A partir do resultado dessa primeira simplificação, o(a) anotador(a) será requisitado(a) a seguir um processo mais rígido de simplificação baseado no manual de simplificação sintática elaborado no escopo do PorSimples (Specia et al., 2008). As simplificações produzidas nessas duas etapas são denominadas, respectivamente: *Simplificação Natural* e *Simplificação Forte*.

Além das etapas do processo de simplificação, há uma etapa prévia na qual o(a) anotador(a) tem a possibilidade de editar/revisar a versão original do texto a ser simplificado. Essa etapa de revisão é realizada com o intuito de corrigir erros simples de digitação e pontuação. Assim, o fluxo de processamento adotado no PorSimples é esquematizado na Figura 1.

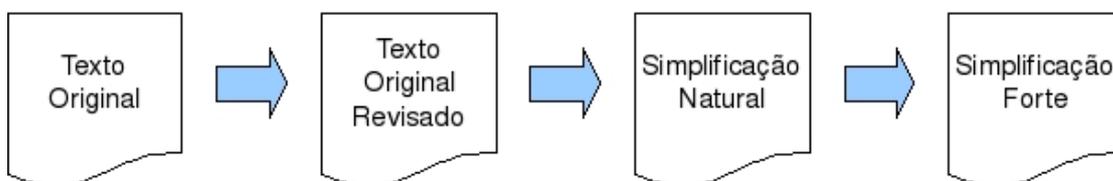


Figura 1. Fluxo do processo de revisão e simplificação no Anotador do PorSimples.

Todo o processamento no Editor de Anotação de Simplificação tem como unidade básica uma *produção*. Uma produção nada mais é do que um texto e seu título. Essa produção é editada e, em seguida, versões simplificadas (natural e forte) são geradas conforme o fluxo de processamento apresentado anteriormente. A tela principal do Editor de Anotação de Simplificação é apresentada na Figura 2.



Figura 2. Tela principal do Editor de Anotação de Simplificação.

Por meio da tela principal do Editor de Anotação de Simplificação, apresentada na Figura 2, o usuário poderá: (1) criar uma nova produção ao clicar no botão *Criar nova produção*, (2) visualizar e revisar uma produção existente (clizando sobre o título da produção desejada, na tabela de produções), (3) simplificar as produções revisadas previamente (clizando em uma das duas colunas referentes às simplificações natural e forte) ou (4) excluir uma produção (clizando no ícone de lixeira presente na última coluna).

A tabela da Figura 2 contém todas as produções criadas previamente e, portanto, armazenadas no banco de dados. Essa tabela está composta por 4 colunas, as quais contém, respectivamente: (1) o título da produção, (2) o bitexto original-natural, (3) o bitexto natural-forte e (4) a opção de remoção de uma produção. A primeira coluna (título) é preenchida para todas as produções armazenadas no banco de dados enquanto as duas seguintes só estarão preenchidas com *links* se o processo de simplificação natural e forte, respectivamente, estiverem habilitados para serem realizados (ou já concluídos) para a produção em questão.

As linhas da tabela da Figura 2 apresentam cores diferentes. Dessas, vale dizer que a linha na cor rosa indica uma produção não habilitada para o processo de simplificação, ou seja, uma produção que foi ignorada no processo de revisão, como explicado a seguir.

Com base no fluxo de processamento adotado no Editor de Anotação de Simplificação do PorSimples, as próximas seções descrevem como cada etapa é realizada por meio da: (seção 2) criação de uma nova produção, (seção 3) edição de uma produção existente, (seção 4) simplificação natural e (seção 5) simplificação forte.

2. Criação de uma nova produção

Para criar uma nova produção, o usuário deverá clicar sobre o botão *Criar nova produção* apresentado na Figura 2. O resultado dessa operação é a tela apresentada na Figura 3.

A imagem mostra a interface web do editor de anotação de simplificação do PorSimples. No topo, há uma barra azul com o logo "PorSimples" e o subtítulo "Simplificando o Português". À direita, o título "Editor de Anotação de Simplificação" é exibido em branco sobre o fundo azul. Abaixo, uma barra verde contém o texto "Criando uma nova produção". O formulário principal é dividido em duas seções: "Título:" com um campo de entrada de texto e "Texto:" com uma área de texto grande e vazia. No canto inferior esquerdo do formulário, há um botão verde com o texto "Criar Produção". No topo da janela, uma barra de menu contém as opções: Arquivo, Editar, Exibir, Histórico, Favoritos, Ferramentas e Ajuda.

Figura 3. Criação de uma nova produção.

Como dito anteriormente, cada produção é formada por um título e um texto. Assim, após inserir o título e o texto nos locais especificados na Figura 3, o usuário deve clicar no botão *Criar Produção* para que a mesma seja armazenada no banco de dados. Nesse caso, uma tela de confirmação será exibida com um *link* para iniciar a simplificação dessa nova produção como descrito na Seção 4.

3. Edição de uma produção

Para editar (ou revisar) uma produção criada previamente e, portanto, armazenada no banco de dados, o usuário deverá clicar sobre o título da produção desejada na tabela apresentada na Figura 2. O resultado dessa operação é a tela apresentada na Figura 4.

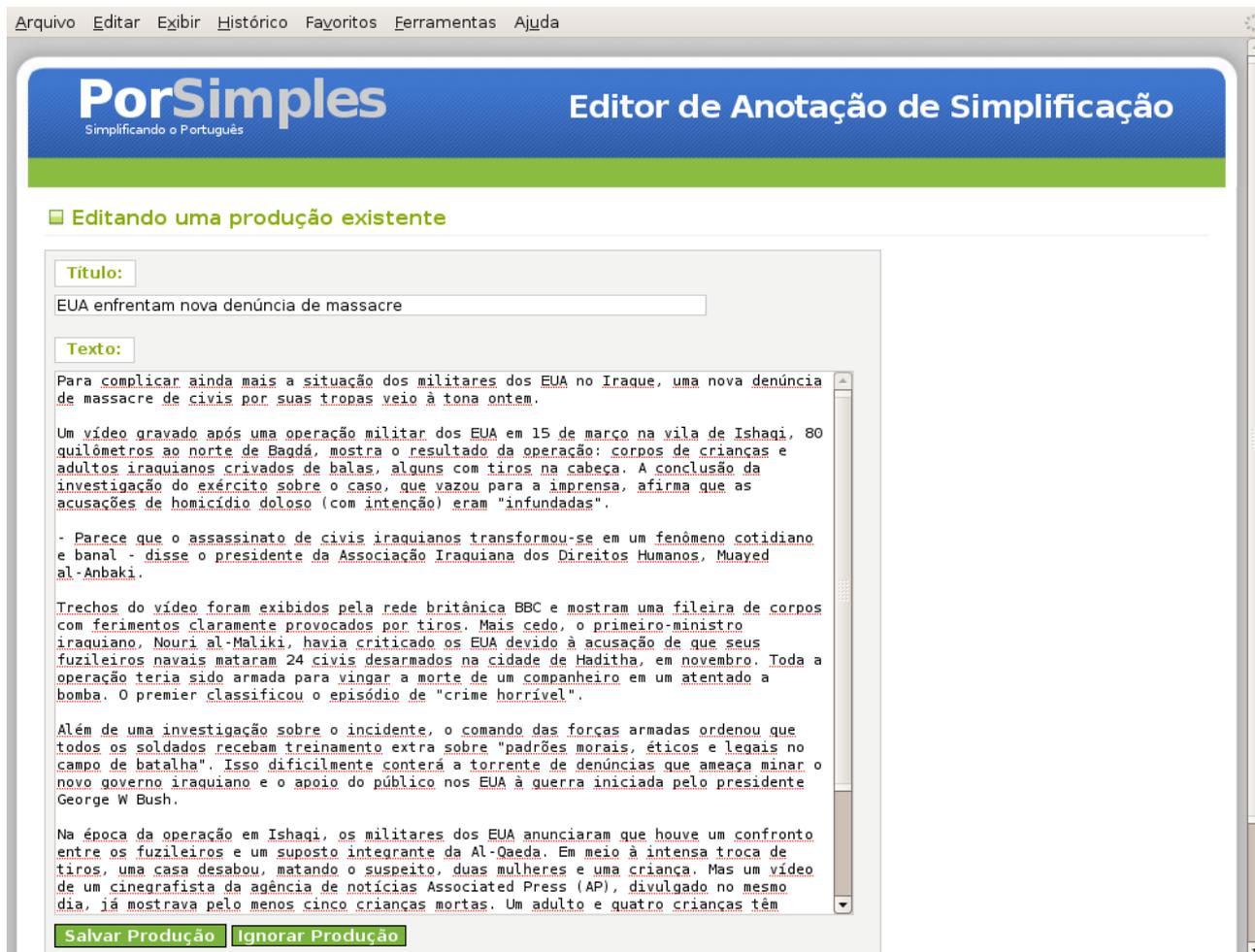


Figura 4. Edição de uma produção existente.

Como apresentado na Figura 4, na tela *Editando uma produção existente*, o usuário poderá alterar o texto inicialmente armazenado no banco de dados.

O usuário pode optar por ignorar a produção, clicando no botão *Ignorar produção*, se observar que a mesma apresenta uma qualidade duvidosa. As produções ignoradas são apresentadas na tabela da tela principal com a cor rosa.

A revisão da produção deverá englobar apenas a correção de erros de pontuação ou digitação. Se o usuário notar que correções mais profundas são necessárias devido à qualidade duvidosa da produção, a mesma deverá ser ignorada.

Se o usuário optar por revisar a produção fazendo apenas as correções permitidas, ele poderá salvá-la clicando no botão *Salvar produção*. Nesse caso, uma tela de confirmação será exibida com um *link* para iniciar a simplificação dessa produção revisada, como descrito na Seção 4.

4. Simplificação Natural

O processo de simplificação no PorSimples tem início com a simplificação natural. Para iniciar a simplificação, o usuário pode proceder de duas maneiras distintas. A primeira é por meio da criação de uma nova produção como explicado na Seção 2; a segunda é por meio da revisão de uma produção já armazenada no banco de dados, como apresentado na Seção 3.

Em ambos os casos, após o usuário selecionar a opção de “Criar/Salvar Produção”, ele será direcionado para a tela apresentada na Figura 5, a qual se refere à primeira produção (aquela cujo título é “Palometas atacam veranistas na fronteira Oeste”) na tabela da Figura 2.

A tela de simplificação natural, assim como a tela de simplificação forte, está dividida em 3 áreas distintas, identificadas pelos números, em preto, 1, 2 e 3:

1. *Texto Original*: apresenta o texto original fornecido pelo usuário no momento da criação ou revisão de uma produção.
2. *Texto Simplificado*: apresenta a versão do texto original com as simplificações sendo feitas pelo usuário.
3. *Operações*: apresenta as operações de simplificação que foram realizadas pelo usuário, sentença-a-sentença.

O usuário deve realizar a simplificação considerando uma sentença original de cada vez, ou seja, sentença-a-sentença. Mais especificamente, para cada sentença original em 1, o usuário deverá escrever em 2 a(s) sentença(s) simplificada(s) correspondente(s). Além disso, o usuário deverá informar as operações que foram realizadas em 1, para que estas operações sejam apresentadas em 3.

A simplificação deve ser realizada sentença-a-sentença e é de fundamental importância que o usuário informe todas as operações de simplificação realizadas em cada sentença.

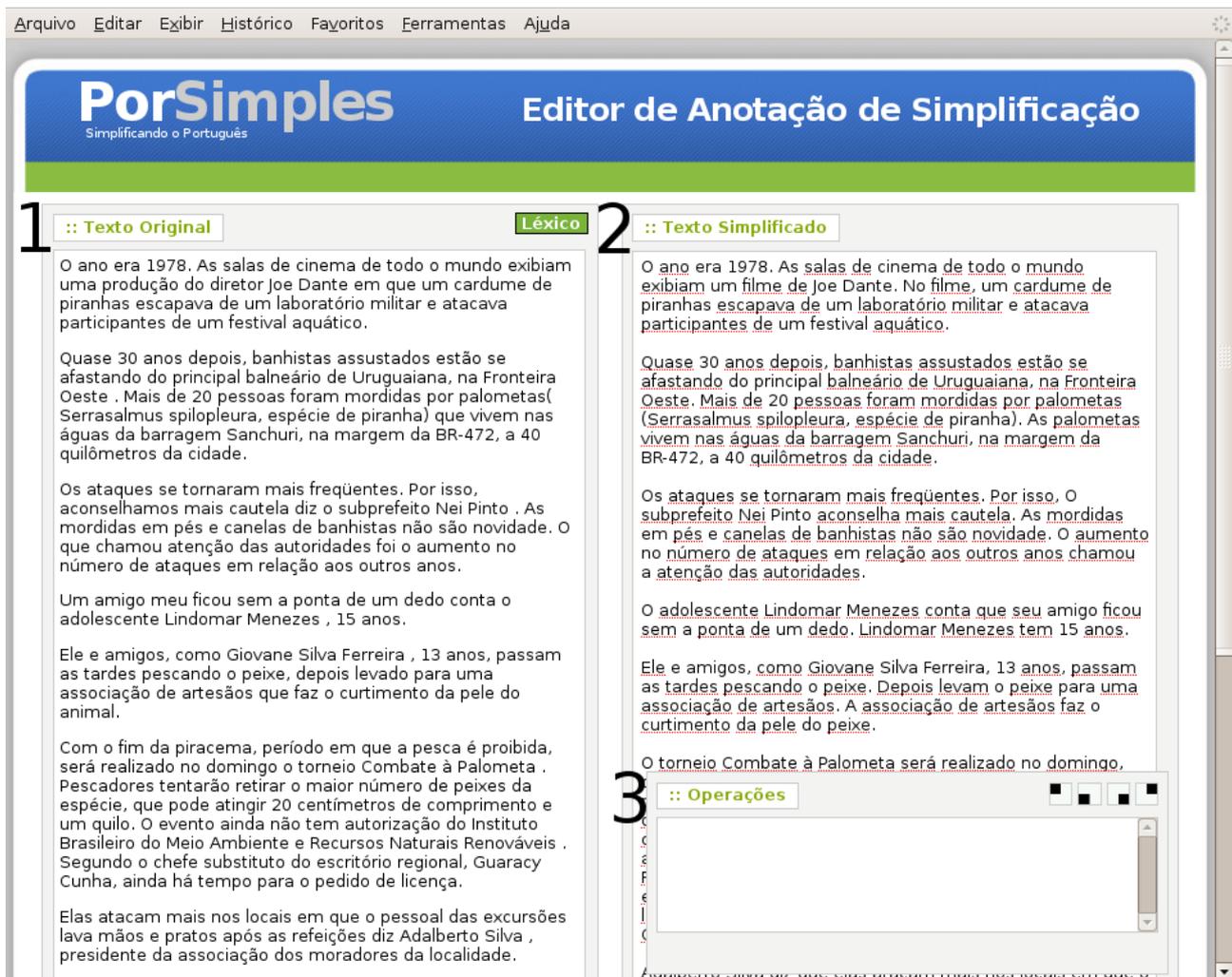


Figura 5. Tela de Simplificação Natural de um texto original

A(s) sentença(s) simplificada(s) correspondente(s) a uma sentença original deve(m) ser introduzida(s) manualmente em 2. As operações de simplificação, por sua vez, são selecionadas a partir de um menu de operações de simplificação acionado ao clicar com o botão esquerdo do mouse sobre uma sentença. O menu de operações de simplificação é apresentado na Figura 6.

Vale lembrar que as operações de simplificação serão inseridas automaticamente na área 3, porém, as alterações necessárias na versão simplificada deverão ser inseridas manualmente na área 2.

Não Simplifica
Operações de reescrita
Fazer reescrita forte
Fazer reescrita simples
Reescrever em SVO
Mudança de voz
Operações de movimentação
Inverter ordem da sentença
Dividir sentença
Unir sentença
Operações de remoção
Remover sentença
Remover parte da sentença

Figura 6. Menu de operações de simplificação.

Operações de Simplificação

As operações de simplificação possíveis e seus significados são:

- **Não simplifica**

A sentença original é mantida intacta, ou seja, sem nenhuma alteração.

- **Operações de reescrita**

- o **Fazer reescrita forte**

A reescrita forte envolve modificações complexas na ordem e/ou no modo de apresentação dos fatos. Por exemplo, Petersen & Ostendorf (2007) apresentam como exemplo de reescrita forte o apresentado abaixo:

Original	<i>The park service says the solution is money.</i>
Simplificada	<i>Why hasn't the National Park Service kept up the park repairs? There is a lack of money.</i>

- o **Fazer reescrita simples**

Ao contrário da reescrita forte, a reescrita simples envolve apenas substituição de palavras ou expressões.

- o **Reescrever em SVO**

A estrutura da sentença é alterada para apresentar o conteúdo na ordem Sujeito-Verbo-Objeto.

- o **Mudança de voz**

A voz da sentença é modificada. A mudança prevista no manual de simplificações (Specia et al., 2008) é a da voz passiva para a ativa.

- **Operações de movimentação**

- o **Inverter ordem da sentença**

A ordem da sentença é invertida, ou seja, duas partes de seu conteúdo trocam de posição.

- o **Dividir sentença**

A sentença original é dividida (dá origem) em mais de uma na versão simplificada, ou seja, o conteúdo da sentença original é apresentado em mais de uma sentença simplificada. O número de sentenças originadas com essa divisão deverá ser informado pelo usuário por meio da tela apresentada na Figura 7.

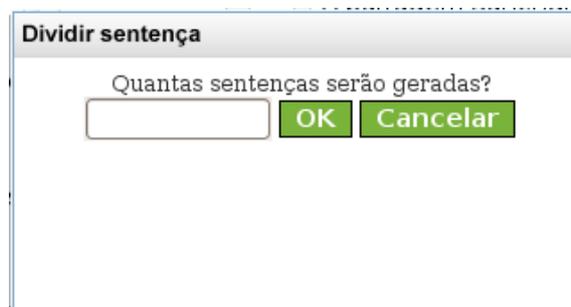


Figura 7. Operação "Dividir sentenças".

- o **Unir sentença**

A sentença original (sobre a qual o usuário clicou) será unida com a sentença anterior (que a precede). **ATENÇÃO:** a união sempre será da sentença selecionada com sua antecedente!

- **Operações de remoção**

- o **Remover sentença**

A sentença original não aparece na versão simplificada.

- o **Remover parte da sentença**

Parte da sentença original não aparece na versão simplificada.

A partir do menu de operações, o usuário deve, então, selecionar a operação de simplificação desejada para a sentença em questão. Assim que o usuário selecionar a operação de simplificação desejada esta será exibida em 3. Esse processo de seleção de operação de simplificação deve ser repetido caso mais de uma operação seja necessária para simplificar uma dada sentença original.

Para ilustrar como essas operações são armazenadas na área 3, vamos supor que as operações “Dividir sentença” e “Fazer reescrita simples” foram realizadas para a simplificação da segunda sentença original. Assim, a área 3 será como apresentado na Figura 8.

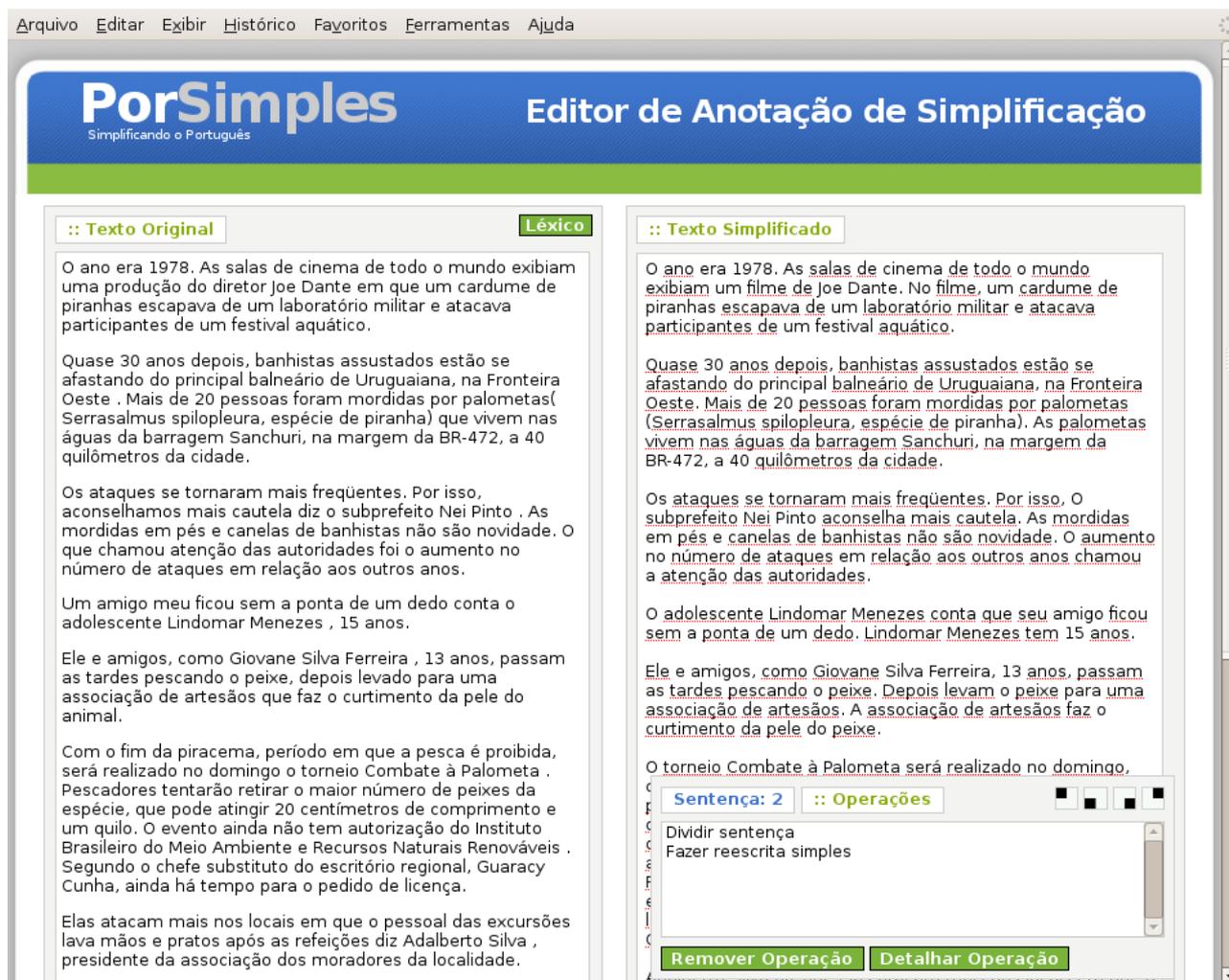


Figura 8. Simplificação natural da segunda sentença original.

A tela de operações de simplificação (área 3) pode ser posicionada nos quatro cantos da tela principal, como especificado pelo desenho no seu canto superior direito apresentado em detalhe na Figura 9. Esse desenho indica que a tela pode ser posicionada, respectivamente, no canto superior esquerdo, inferior esquerdo, inferior direito ou superior direito.



Figura 9. Posicionamento da tela de operações.

A tela de operações também conta com dois botões, um para remover uma operação selecionada previamente (*Remover Operação*) e outro para detalhar uma

operação (*Detalhar Operação*), por exemplo, por meio da inserção de algum comentário. Ao clicar no botão de remoção da operação de simplificação, a mesma será removida da lista de operações efetuadas para aquela sentença e cabe ao usuário desfazer qualquer alteração que tenha feito no texto simplificado relacionada à operação removida.

Ao clicar no botão de detalhamento da operação, por sua vez, a tela apresentada na Figura 10 será mostrada.

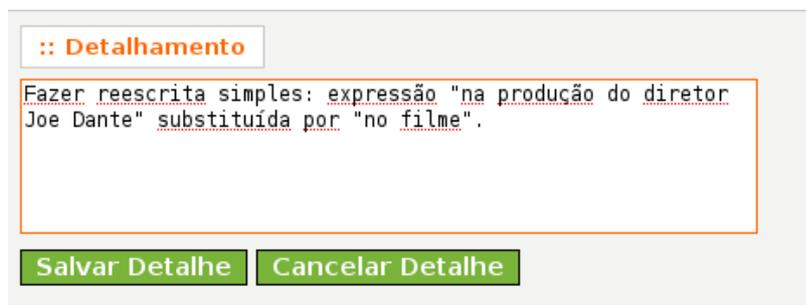


Figura 10. Detalhamento de uma operação de simplificação.

Se algum detalhamento já estiver armazenado para essa operação, no banco de dados, o mesmo será mostrado na tela, caso contrário, a tela aparecerá com o conteúdo em branco. Em ambos os casos, o usuário poderá editar o detalhamento e salvá-lo (por meio do botão *Salvar Detalhe*) ou cancelar o detalhamento (por meio do botão *Cancelar Detalhe*). Por exemplo, na Figura 10, há um detalhamento armazenado no banco de dados para a operação de reescrita simples para a terceira sentença original.

Na versão atual do Editor de Anotação de Simplificação, o texto do detalhamento de uma operação deverá mencionar, necessariamente, o nome da operação a qual esse detalhamento se refere.

O Léxico

No processo de Simplificação Natural, além das operações de simplificação apresentadas no menu, o usuário também conta com um recurso extra: o “Léxico”. Para dispor deste recurso, o usuário deve clicar no botão “Léxico” posicionado no canto superior direito da área 1, como mostra a Figura 11.

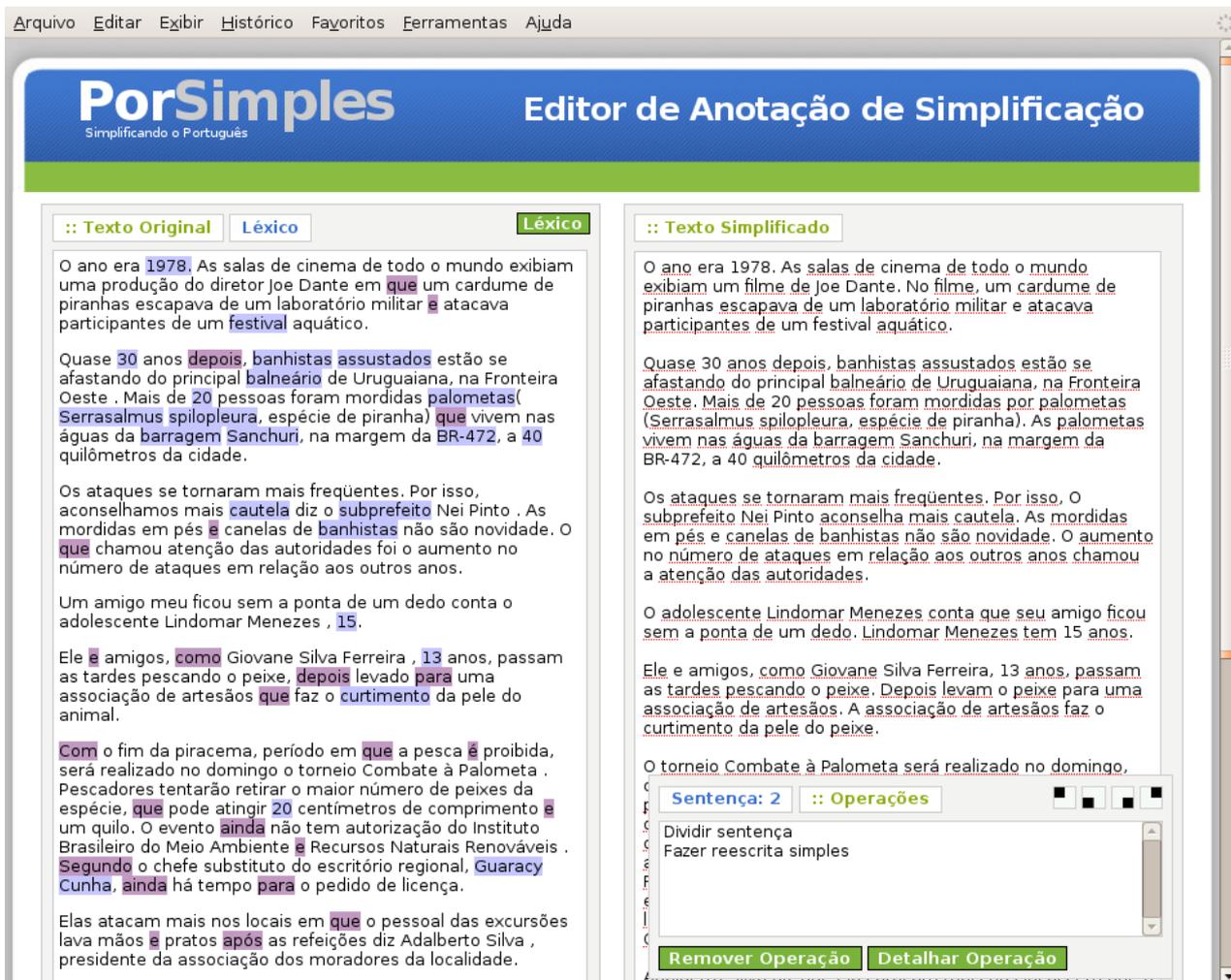


Figura 11. Uso do "Léxico" para identificação de marcadores discursivos e palavras complexas

Ao habilitar o uso do Léxico, as palavras consideradas complexas ou os marcadores discursivos ambíguos ou complexos são realçados no texto, como mostra a Figura 11. Para que essa funcionalidade seja possível, duas listas são usadas: uma lista de palavras simples e outra de marcadores discursivos. A lista de palavras simples foi criada a partir de 3 recursos: Dicionário Ilustrado Infantil (Biderman, 2005); Lista de palavras do Suplemento "Para seu Filho Ler" do Jornal Zero Hora; Lista de palavras concretas construída em (Janczura, 2007). A lista de marcadores discursivos, por sua vez, foi construída com base no trabalho de (Pardo, 2005).

As palavras que não constam na lista de palavras simples são consideradas palavras complexas e realçadas no texto em um tom azulado (violeta), por exemplo a palavra *festival* na segunda sentença do texto da Figura 11. Para as palavras complexas, é sugerida a operação de substituição da palavra por outra mais simples como mostra a tela da Figura 12 acionada ao passar o mouse sobre uma palavra marcada como complexa.

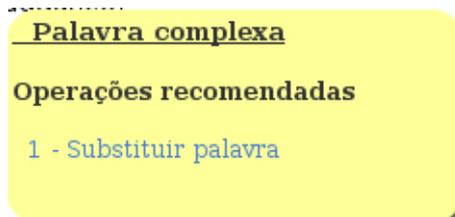


Figura 12. Substituição de uma palavra complexa.

Os marcadores discursivos ambíguos ou complexos, por sua vez, são realçados em um tom roxo claro como, por exemplo, o marcador *que* na segunda sentença original da Figura 11. Para os marcadores, em casos previstos no Manual de Simplificação (Specia et al., 2008), a ferramenta sugere possíveis trocas por um dos marcadores citados no manual ou por outro especificado pelo usuário, como indicado na Figura 13. Neste caso, as opções de substituição são apresentadas como uma lista enumerada sempre com o texto inicial "Substituir marcador por". O que segue este texto padrão é a opção sugerida para a substituição, apresentada entre aspas.

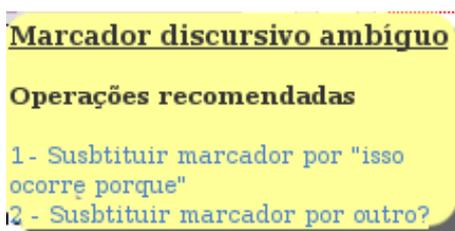


Figura 13. Operações para marcadores discursivos.

Como apresentado na Figura 13, o primeiro item é uma sugestão fornecida de acordo com o Manual de Simplificação (*isso ocorre porque*) enquanto o segundo item é uma opção para que o usuário escolha um outro marcador não especificado. No segundo caso, uma mensagem pedindo que o usuário digite o novo marcador será apresentada.

É importante lembrar que as operações de substituição lexical, realizadas como descrito anteriormente, serão inseridas automaticamente na área 3, porém, a substituição de palavras e marcadores discursivos na versão simplificada deverá ser inserida manualmente na área 2.

Finalizando uma Simplificação Natural

Após a simplificação de todas as sentenças originais (na área 1) por meio da realização das modificações no texto simplificado (na área 2) e da indicação de todas as operações de simplificação efetuadas (na área 3), o botão *Salvar Simplificação Natural* deverá ser acionado para armazenar o bitexto Original-Natural, juntamente com todas as informações correspondentes, no banco de dados.

Essa ação levará o usuário novamente para a tela principal do Editor de Anotação de Simplificação.

A próxima seção descreve o processo de criação da Simplificação Forte, o qual deverá ter como base (entrada) uma Simplificação Natural. Portanto, apenas as produções já simplificadas “naturalmente” poderão ser simplificadas por meio do processo “forte”.

5. Simplificação Forte

Para ilustrar o processo de criação de uma Simplificação Forte vamos supor que o usuário já tenha realizado a Simplificação Natural da primeira produção (aquela cujo título é “Palometas atacam veranistas na fronteira Oeste”) como descrito anteriormente.

A tela da Simplificação Forte é muito semelhante à da Simplificação Natural, apresentada anteriormente, só que, nesse caso, para a geração da Simplificação Forte, como mostra a Figura 14.

:: Texto Natural **Léxico** **Sintático**

O ano era 1978. As salas de cinema de todo o mundo exibiam um filme de Joe Dante . No filme, um cardume de piranhas escapava de um laboratório militar e atacava participantes de um festival aquático.

Quase 30 anos depois, banhistas assustados estão se afastando do principal balneário de Uruguaiana, na Fronteira Oeste . Mais de 20 pessoas foram mordidas por palometas(Serrasalmus spilopleura, espécie de piranha) . As palometas vivem nas águas da barragem Sanchuri, na margem da BR-472, a 40 quilômetros da cidade.

Os ataques se tornaram mais freqüentes. Por isso, O subprefeito Nei Pinto aconselha mais cautela. As mordidas em pés e canelas de banhistas não são novidade. O aumento no número de ataques em relação aos outros anos chamou a atenção das autoridades.

O adolescente Lindomar Menezes conta que seu amigo ficou sem a ponta de um dedo. Lindomar Menezes tem 15 anos.

Ele e amigos, como Giovane Silva Ferreira , 13 anos, passam as tardes pescando o peixe. Depois levam o peixe para uma associação de artesãos. A associação de artesãos faz o curtimento da pele do peixe.

O torneio Combate à Palometa será realizado no domingo, com o fim da piracema. A piracema é o período em que a pesca é proibida. Pescadores tentarão retirar o maior número de peixes da espécie. A espécie pode atingir 20 centímetros de comprimento e um quilo. O evento ainda não tem autorização do Instituto Brasileiro do Meio Ambiente e Recursos Naturais Renováveis . O chefe substituto do escritório regional diz que ainda há tempo para o pedido de licença. O chefe substituto do escritório regional é Guaracy Cunha .

Adalberto Silva diz que elas atacam mais nos locais em que o pessoal das excursões lava mãos e prato após as refeições. Adalberto Silva é presidente da associação dos moradores da

:: Texto Simplificado

O ano era 1978. As salas de cinema de todo o mundo exibiam um filme de Joe Dante. No filme, um cardume de piranhas escapava de um laboratório militar. O cardume de piranhas atacava participantes de um festival aquático.

Quase 30 anos depois, banhistas assustados estão se afastando do principal balneário de Uruguaiana, na Fronteira Oeste. Palometas morderam mais de 20 pessoas. Palometas (Serrasalmus spilopleura) são uma espécie de piranha. As palometas vivem nas águas da barragem Sanchuri, na margem da BR 472, a 40 quilômetros da cidade.

Os ataques se tornaram mais freqüentes. Por isso, o subprefeito Nei Pinto aconselha mais cautela. As mordidas em pés e canelas de banhistas não são novidade. O aumento no número de ataques em relação aos outros anos chamou a atenção das autoridades.

O adolescente Lindomar Menezes conta que seu amigo ficou sem a ponta de um dedo. Lindomar Menezes tem 15 anos.

Ele e amigos passam as tardes pescando o peixe. Depois levam o peixe para uma associação de artesãos. A associação de artesãos curte a pele do peixe.

O torneio Combate à Palometa será realizado no domingo, no fim da piracema. A pesca é proibida na piracema. Pescadores

:: Operações

Figura 14. Simplificação Forte de uma Simplificação Natural.

Para a realização da Simplificação Forte, o usuário dispõe das mesmas operações de simplificação e o recurso Léxico disponíveis para a Simplificação Natural. Além desses recursos, na Simplificação Forte o usuário dispõe de outro recurso: o Sintático.

O Sintático

O recurso Sintático, acessível ao clicar sobre o botão no canto superior direito da área 1 da Figura 15, é responsável por realçar características sintáticas que, de acordo com as especificações do Manual de Simplificação (Specia et al., 2008), merecem a atenção do(a) anotador(a) no processo de Simplificação Forte.

Essas informações sintáticas são derivadas com o auxílio do *parser* PALAVRAS (Bick, 2000) e, entre elas, podem ser citadas as apresentadas na Figura 15.

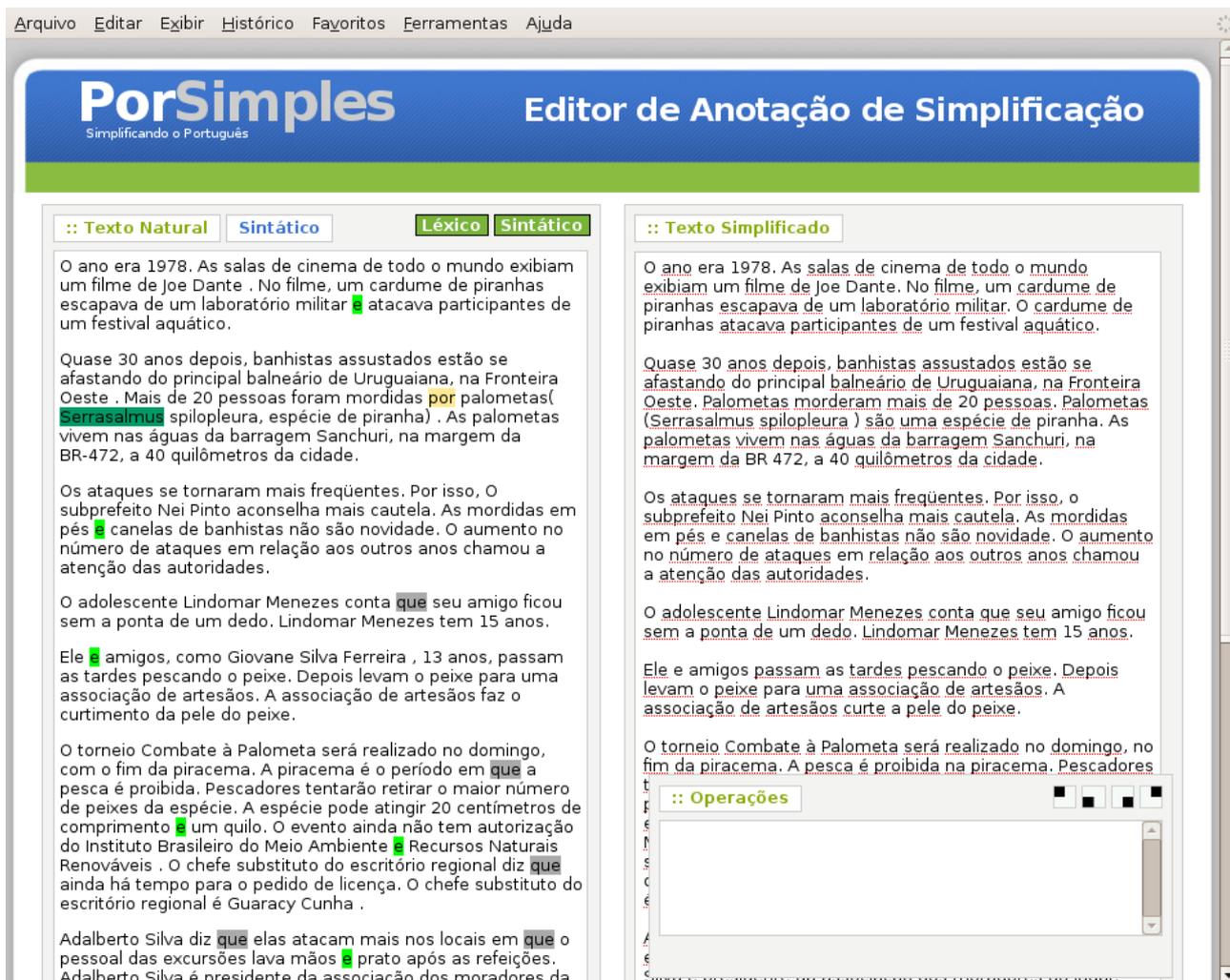


Figura 15. Uso do "Sintático" para a identificação de características sintáticas relevantes.

De maneira semelhante ao recurso "Léxico", um menu de opções acompanha cada característica sintática realçada no texto. Esse menu apresenta a operação sugerida no manual para o tipo de característica em questão. Por exemplo, a Figura 16 apresenta a operação "Dividir sentença" sugerida para a oração coordenada presente na segunda sentença original. Ao selecionar essa opção a mesma tela da Figura 7 será apresentada já que se trata da mesma operação de divisão de sentenças mencionada anteriormente.

Na Simplificação Forte o usuário deve seguir as recomendações do Manual de Simplificação, as quais são sugeridas para cada característica sintática mencionada no Manual.

Assim como no processo de Simplificação Natural, na Simplificação Forte, as operações de simplificação léxica e sintática devem ser selecionadas para cada sentença original na área 1, modificando o texto na área 2 e indicando as operações realizadas, na área 3.

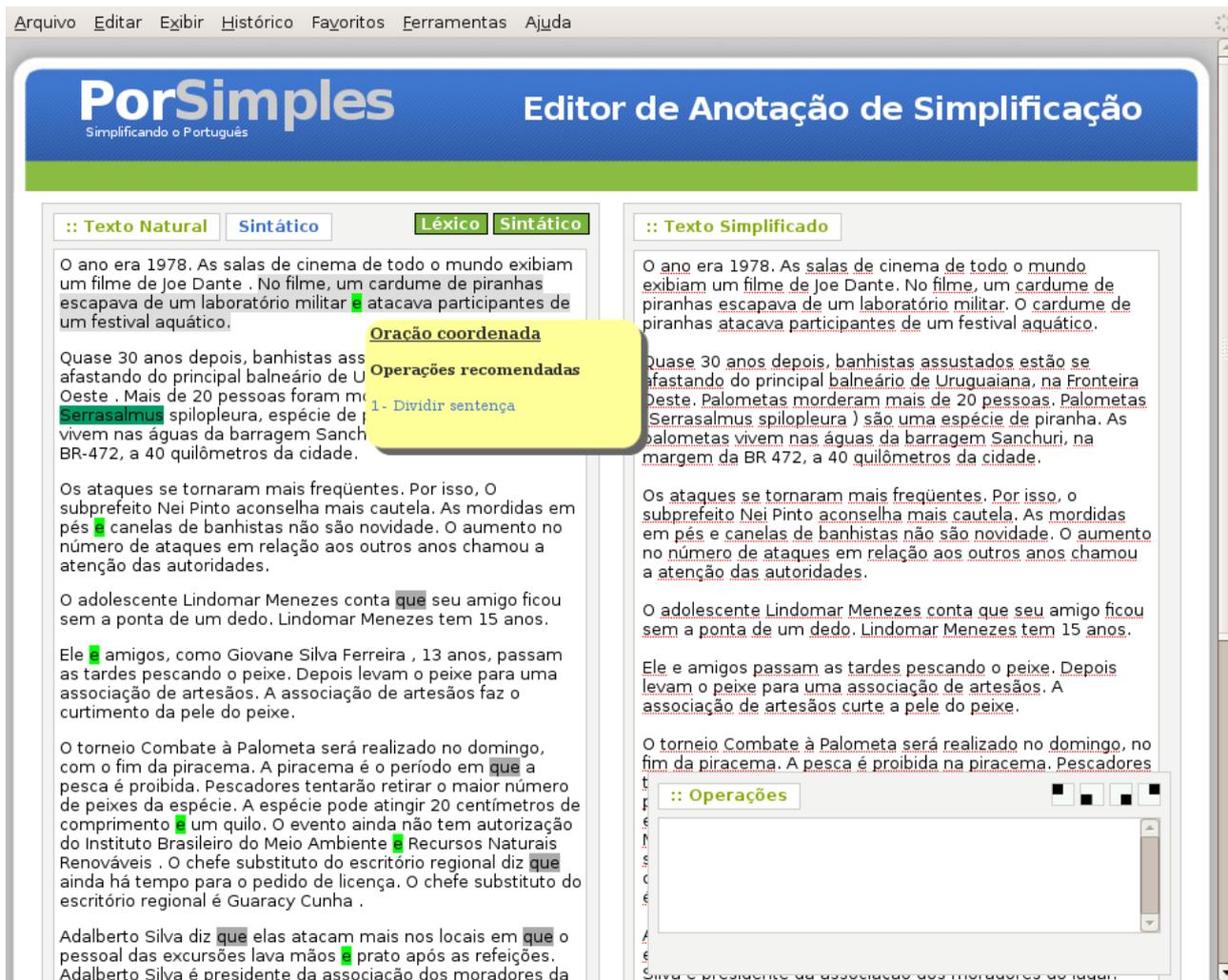


Figura 16. Operações de simplificação possíveis para a segunda sentença.

Finalizando uma Simplificação Forte

Ao final do processo de simplificação, o botão *Salvar Simplificação Forte* deve ser acionado para salvar todas as informações referentes à Simplificação Forte no banco de dados. Essa ação finaliza o processo de simplificação da produção como ilustra a mensagem da Figura 17.



Figura 17. Tela final do processo de Simplificação no PorSimples.

6. Referências Bibliográficas

- Bick, E. (2000). *The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. PhD thesis – Aarhus University. Aarhus, Denmark: Aarhus University Press, November 2000.
- Biderman, M. T. C. (2005). *Dicionário Ilustrado de Português*. São Paulo, Editora Ática. 1ª Edição. São Paulo: Ática.
- Janczura, G. A., Castilho, G. M., Rocha, N. O. (2007). *Normas de concretude para 909 palavras da língua portuguesa*. *Psic.: Teor. e Pesq.* [online]., vol. 23, p. 195-204.
- Pardo, T. A. S. (2005). *Métodos para Análise Discursiva Automática*. Tese de Doutorado. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo. São Carlos-SP, Junho, 211p.
- Petersen, S. E.; Ostendorf, M. (2007). Text Simplification for Language Learners: A Corpus Analysis. In: *Workshop Speech and Language Technology for Education*. Pennsylvania, USA. October. p. 69-72.
- Specia, L.; Aluisio, S.M.; Pardo, T.A.S. (2008). *Manual de Simplificação Sintática para o Português*. Technical Report NILC-TR-08-06. São Carlos-SP.