

TBTA Tutorial Lesson 5: Feature Copying Rules and Find/Replace Rules

(video: [TBTA-5A Feature Copying Rules](#))

(video: [TBTA-5B Find/Replace Rules](#))

5.0 Introduction

In the previous lesson you wrote several spellout rules and modified two phrase structure rules. Hopefully you're now fairly familiar with those two types of rules. In that lesson you also generated some very simple sentences, saved the results, and then compared the saved text with the newly generated text.

In this lesson we're going to look at two more types of rules in TBTA's synthesizing grammar. The first type is called Feature Copying rules, and the second type is called Find/Replace rules. Both of these rule types are very easy to use.

5.1 Feature Copying Rules

In lesson 3 you entered a lexical form for your target verbs called Third Singular Present. This form is necessary because English verbs agree in person and number with their subject nouns. When the subject noun is 3rd person singular and the verb is present tense (e.g., *He walks.*), the verb is marked with the suffix *-s*. We now need a spellout rule in the grammar that will choose the Third Singular Present form from the lexicon when a verb is present tense, and the subject noun is 3rd person singular. Use Grammar Wizard to write a rule that will select the Third Singular Present form for verbs from the lexicon; to do this you'll use the fifth template. When you see the rule, click the Features button. When you look at the features for verbs, you'll see that the Person and Number of the subject noun aren't available in the features dialog. This is because Person and Number are features on nouns rather than verbs. So before we can write a spellout rule for verbs that will select the Third Singular Present form from the lexicon, we need to write a feature copying rule that will copy the Person and Number features from the subject noun to the verb. That is the purpose of feature copying rules; they copy features from one constituent in a sentence to another constituent in the same sentence. After the relevant features have been copied, the spellout rules for verbs will be able to access those copied features and perform the appropriate actions. So cancel the spellout rule that you're currently writing and let's use Grammar Wizard to write a feature copying rule that will copy the value of the Person and Number features from the subject noun to the verb. Then we'll be able to write a spellout rule which will select the Third Person Singular form from the lexicon when the subject noun is 3rd person singular and the verb tense is present.

Use Grammar Wizard and look at the templates. You'll see that the first template is used for feature copying rules, so double click on the first template.

Then double click on 'Number from Nouns'. We want this rule to copy both Person and Number, but in Grammar Wizard we can only specify one feature to copy. Then double click on verbs because we want to copy these features to verbs. You'll then see the feature copying rule dialog. As always the rule needs to be put in a group, so enter a group called **Subject Agreement**. Below the rule's name and the two rows of buttons you see a structure. This structure shows that this rule will copy features from a noun to a verb. Since we want the features to be copied from the subject noun, we need to specify the features of the NP. We still haven't written rules that generate grammatical relations from semantic roles, so for this rule we'll copy the Person and Number features from nouns in Agent NPs. To specify that the NP must be an Agent NP, click on the Features button. You'll see that it remains depressed indicating that you're ready to specify features. Then click on the opening NP phrase boundary in the structure and select Most Agent-like in the Semantic Role column. Close the features dialog and rest the cursor on the opening NP bracket. You'll see that the Semantic Role is Most Agent-like. You're now ready to specify which features should be copied.

In the dropdown labeled 'Feature to Copy' you see 'Number'; this was entered by Grammar Wizard. We need to specify the name of this copied feature so enter **Number of Subject** into the window below the dropdown. We'll leave the default value set to Source Unavailable. We're now ready to make this rule copy a second feature so click the scroll button in the upper left corner of the frame 'Features to be Copied' so that you see '2/1'. Open the Feature to Copy dropdown and you'll see all of the features associated with nouns. Select **Person** because this is the second feature we want this rule to copy. Then enter a name for this copied feature such as **Person of Subject**. Assign this rule to the topics Number and Tense. Since this rule copies two features from the subject nominals to the verb, let's change the rule's name to "Copy Person and Number from Subject Nouns to Verbs." You've now completed this feature copying rule; it should look like the one shown below in Figure 5-1.

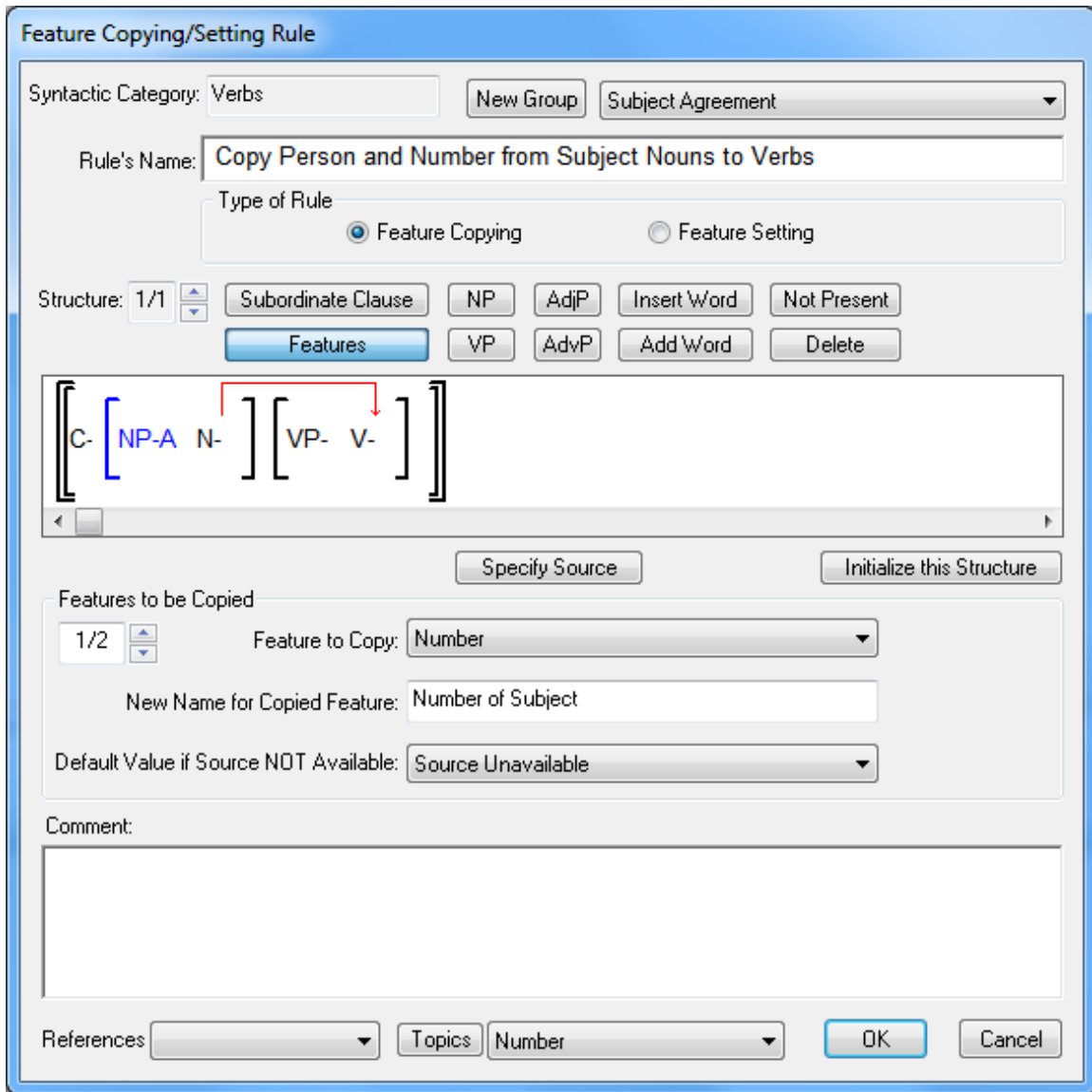


Figure 5-1. Feature Copying Rule

To test this rule, go to Verbs 1:2. The window below the grammar tree indicates that this verse will generate *John walks*. Right click on the concept JOHN and map that concept to the target word *John*. Then click the Generate button. After the verse has been generated, if you rest the cursor on the features under WALK, the popup will show you the features in the semantic representation, then a dashed line and the feature that you defined for verbs in the target lexicon. Then you'll see another dashed line and the two features that were copied from the NP-Agent noun to the verb. You'll see that the last two features on the verb have been copied from the NP-Agent noun and are 3rd Person and Singular. This popup is shown below in Figure 5-2. Now that we know the feature copying rule is working properly, let's use this information to write a spellout rule for verbs that will select the Third Person Singular form that we entered into the target lexicon.

Verb
Semantic Complexity Level = Molecule
Lexical Sense = A
Time = Present
Aspect = Unmarked
Mood = Indicative
Reflexivity = Not Applicable
Polarity = Affirmative
Target Tense = Unspecified
Target Aspect = Unspecified
Target Mood = Unspecified
----- User Defined Lexical Features for Verbs -----
----- Copied Features for Verbs -----
Number of Subject = Singular (copied from a Noun)
Person of Subject = Third (copied from a Noun)

Figure 5-2. Popup with Semantic Features, Lexical Features and Copied Features

Use Grammar Wizard again to write a rule that will select the Third Singular Present form from the lexicon. Now when you click on the Features button in that rule, scroll to the far right. You'll see two new columns with a blue background. The names of those columns are 'Number of Subject' and 'Person of Subject', just as you specified in the feature copying rule. Use this dialog to specify that verb Time must be Present, Person of Subject is Third, and Number of Subject is Singular. Assign this rule to the grammar topics Number, Tense and Grammatical Relations. Your rule should now look like that shown in Figure 5-3.

Spellout Rule

Syntactic Category: Verbs New Group Tense

Rule's Name: Select the Third Singular Present form for verbs

Status On

Type of Rule Simple Table Morphophonemic Form Selection Phrase Builder

Structures

Base Form: Third Singular Present

Features

- Verb Time = Present
- Verb Number of Subject = Singular
- Verb Person of Subject = Third

Comment:

Verbs 1:2 John walks.

References: Topics Grammatical Relations OK Cancel

Figure 5-3. Spellout Rule to select Third Singular Present form for Verbs

Test this rule by clicking the Generate button. You should see *The John walks*. The agreement and spellout rules are now working correctly, but the spellout rule for nouns that inserts the Article *the* needs to be revised so that it only applies to common nouns, not proper nouns. Revise that rule accordingly and you should then generate *John walks*. After you've generated the desired result, click the Save Verse button.

Rest the cursor on the word *walks* in the large window, and you'll see a popup showing the history of that word. You'll see that the features were changed by the feature copying rule that you wrote, and then *walk* was changed to *walks* by the form selection spellout rule that you wrote. After you've generated a verse, you can always inspect the history of each word by resting your cursor on that word.

When a feature copying rule copies the Number feature from Nouns, TBTA knows that it must find all of the relevant nouns in the sentence. In the rule that

you just wrote, you specified that the noun must be in the Agent NP. So TBTA will find all of the agent NPs in the sentence and add their number values together (e.g., Singular plus Singular is Dual, Singular plus Dual is Trial, Dual plus Dual is Quadrial, etc.). For example, go to Noun Phrases 1:2. The window immediately below the grammar tree indicates that this verse will generate *John and Mary walked*. Push the Generate button and then look at the Number of Subject feature on the verb by resting the cursor on the verb's features. You'll see that the Number of Subject feature value is Dual. In English we only specify singular and plural; duals, trials and quadrials are all treated as plurals. In a subsequent lesson we'll learn how to convert duals, trials and quadrials to plurals. All you need to know now is that TBTA recognizes that when Number is copied from nouns, all of the relevant nouns will be found and their number values will be added together.

Open the Feature Copying Rules node in the grammar tree. Then open the Verbs node. You'll see that your new group has been inserted under verbs. Open the group node and double click on the rule's name. You'll then see your feature copying rule. Feature copying rules are always put into the destination's syntactic category in the grammar tree. Since this rule copies features from nouns to verbs, the destination syntactic category is verb. If you use Grammar Wizard to write a feature copying rule that copies a feature from verbs to clauses, you would find that rule under the clause node. Notice that one feature copying rule may copy up to ten features from the source to the destination. If you rest your cursor on the red line going from the noun to the verb in the rule's structure, you'll see a popup showing you all of the features that will be copied by this particular rule.

5.2 Find/Replace Rules

We're now ready to build a Find/Replace rule. These rules are the easiest to use and understand. These rules work just like the Find/Replace option in all word processors. The Find/Replace rules are executed after all of the other types of rules have finished being executed. For each of these rules you specify an input string and an output string. The generated text is then searched for the rule's input string. If the input string is found, that text is removed from the target text and the output string is inserted. When TBTA built your database, it included one Find/Replace rule. Go to the grammar tree and open the Find/Replace Rules node. Find/Replace rules operate on strings rather than syntactic categories, so there are no syntactic categories in the Find/Replace rule node. Instead you'll see a group call Punctuation. Open the group node and you'll see a rule named "Change ' .' to '.'". This rule searches the generated text for a space followed by a period. When it finds this combination, it replaces it with a period. Turn this rule off and generate any verse. In the upper window that shows the final target text, you'll see that there's a space between the last word and the period. Turn the rule back on and generate the verse again. Now you'll see that the period immediately follows the final word. The Find/Replace rules

don't operate on the text in the large window; they only operate on the text in the small window at the top of the screen that shows the final target text.

In the exercises for lesson 4 you wrote a spellout rule for clauses that inserts a question mark into all questions. Go to Clauses 1:91 and make sure that your rule still operates properly; in the large window you should see a “?” inserted between the VP and the patient NP. Remember that the Patient NP is inserted at the end of the sentence because the phrase structure rule for clauses hasn't yet been told where to put it. Modify your clause phrase structure rule so that the question mark follows patient NPs and precedes the period. To do this, you'll need to insert a NP into the rule, and then specify that it is a Patient NP. You'll also need to insert a “period” into the rule. Note that when you click the ‘Insert New Constituent into Rule’ button on the phrase structure rule dialog, the Constituent dropdown in the new dialog includes an option labeled ‘period’. Your revised clause phrase structure rule should now look like that shown in Figure 5-4.

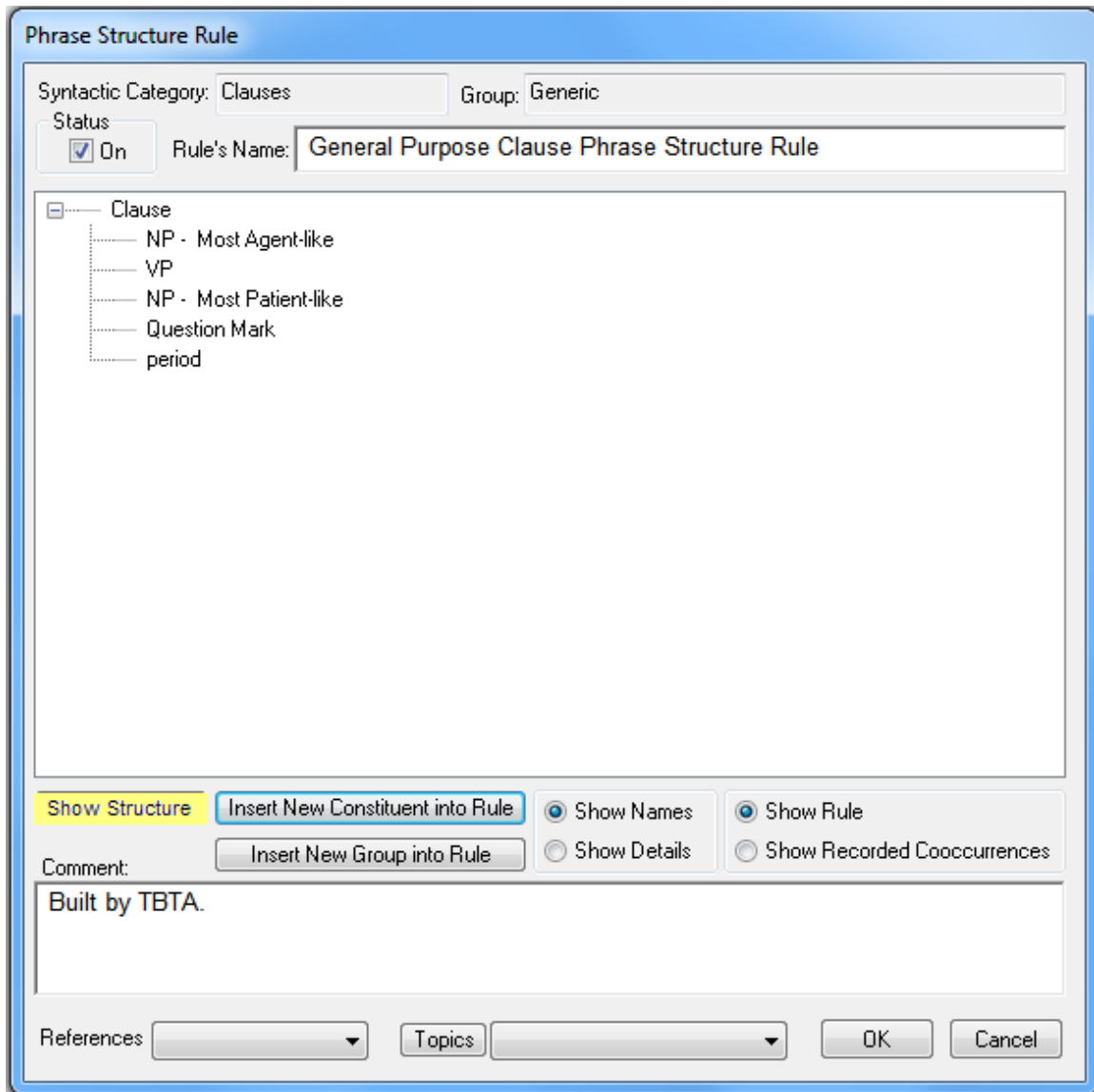


Figure 5-4. Revised Phrase Structure Rule for Clauses

Go to Clauses 1:91 and link the concepts to target words. When linking PERSON, you'll need to enter a new noun stem into the lexicon. To do this, open the Mapping dialog and click the New Target Stem button at the bottom. You'll be prompted to enter a new stem, so enter **person**. You'll then be asked to set the features for this new stem; specify Masculine and Common. Click the OK button and you'll be asked to confirm the plural form. The form *persons* is incorrect so right click on that cell and enter **people**. The gloss will be provided for you so click the OK button. You'll then see this new stem automatically linked to PERSON so you may close the mapping dialog. After you've linked all the concepts in this verse to target words, click the Generate button. The generated text should be "*Person read a book ?*" Note that there's a space preceding the question mark. This is because punctuation marks are treated as if they were words, and words are always separated by spaces. To eliminate the space

before the question mark, let's use Grammar Wizard to write another Find/Replace rule.

Open Grammar Wizard and double click on the last template. For the input string, enter " ?" (a space followed by a question mark). For the output string, enter just "?". Then push the Tab key. Put the new rule into the Punctuation group. Your rule should now look like that shown in Figure 5-5.

The image shows a software dialog box titled "Find / Replace Rule". At the top, there is a "New Group" button and a dropdown menu currently set to "Punctuation". Below this is a "Status" section with a checked checkbox and the text "On". The main area contains four input fields: "Rule's Name:" with the text "Change ' ?' to '?'", "Input string:" with " ?", "Output string:" with "?", and "Comment:" with "Clauses 1:91 Who read a book?". At the bottom, there are two more dropdown menus labeled "References" and "Topics", and two buttons labeled "OK" and "Cancel".

Figure 5-5. Find/Replace rule for eliminating the space before a ?

Save your rule and click the Generate button. Now the question mark should immediately follow *book* and the period immediately follows the question mark. Go ahead and write another Find/Replace rule that will change "?. " to "?". Note that we could have written just one rule that changes " ?." to "?". However, punctuation becomes more complex when you work with periods, question marks

and exclamation marks inside quotation marks. So it's a good idea to have a Find/Replace rule that eliminates the space before each type of punctuation mark. For example, if you write a spellout rule that enters a comma into various types of adverbial clauses, you'll need a Find/Replace rule that will convert 'space comma' to a comma. You can also use Find/Replace rules to generate contractions (e.g., *do not* becomes *don't*, etc.). As you can see, these Find/Replace rules perform very minor cleanup operations at the end of the generation process.

5.3 Capitalizing the First Letter of each Sentence

Let's perform one more minor cleanup before finishing this lesson. Not all languages use alphabets that have capital letters, but English does and we want the first letter of each sentence to be capitalized. To do this, go to Ontology and Lexicon, Lexicon, Nouns. Click on the button labeled Setup, and then click the button labeled 'Target Alphabetizing Character Sequence'. You'll then see a dialog that contains the alphabet you're using. In the lower center of that dialog is a button labeled 'Capital Letters'. Click that button. You'll then see a table with just two columns. The first column contains all of the characters in your alphabet. In the second column you see the characters that are the capitals of the characters in the first column. This table was filled in for you when you created this database. At the bottom of this dialog you see a small window for the sentence final characters. A period has automatically been entered for you, but you need to enter any other characters that signal the end of a sentence in your language. For example, question marks signal the end of a sentence in English, so enter a ? into that window after the period (don't put a space between the period and question mark). Your capitalization dialog should look like Figure 5-6 shown below. Close the dialog by clicking the OK button.

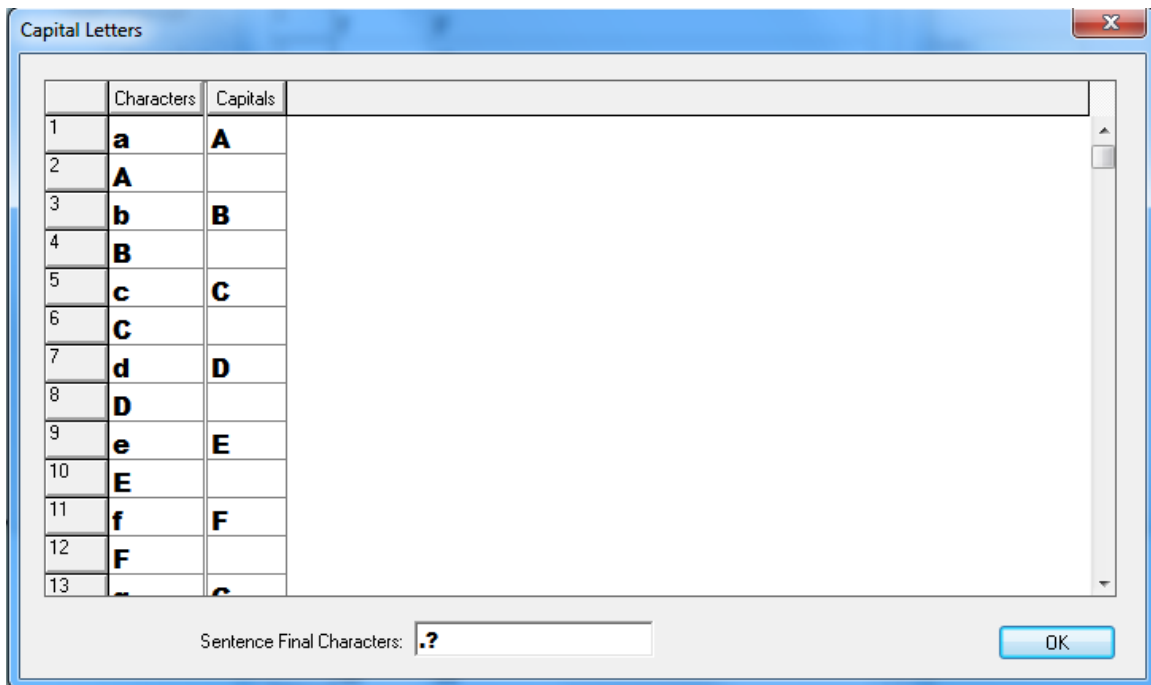


Figure 5-6. The Capitalization Dialog

5.4 Conclusion

In this lesson you learned how to use two more types of rules in TBTA's synthesizing grammar. You wrote a feature copying rule for verbs that copied two features from nouns in agent NPs to the verb. You also wrote several Find/Replace rules that cleaned up the punctuation at the end of each sentence.