TBTA Tutorial Lesson 3: The Target Lexicon

(video: TBTA-3A Lexicon Overview)
(video: TBTA-3B Lexical Spellout Rules)

(video: TBTA-3C Lexical Table Spellout Rules)

3.0 Introduction

This tutorial will introduce you to the target lexicon. You will enter some noun and verb stems into the target lexicon, and you will specify features for those stems. You will also write several rules that will generate various forms from the stems. Throughout these tutorials you will be working with English, but you should be thinking about how the principles described in these lessons will apply to your particular target language.

3.1 Beginning a new Language Project

Let's begin by creating a new database for these tutorials. To create this database, select Utilities in the main menu, then go to Select / Add a Target Language. Click on the Add a New Target Language button in the upper left corner. Enter a name for this language such as My English. You'll then see the system begin building Case Frame Adjustment rules for your transfer grammar. Those rules will be described in a subsequent tutorial. You'll then be prompted to select a font for this language. You may select any font that you like, but choose a font other than Arial so that you'll be able to distinguish it from the standard font used throughout this program. A font such as Arial Black is a good choice. After selecting a font, answer Yes when asked if you want the Roman alphabet to be entered. You'll then be prompted to specify which of the predefined phonetic features are relevant to your language. We'll need these phonetic features when we're developing morphophonemic rules, so go ahead and select all of them by checking the boxes at the left of each row. Then click the OK button to close the phonetic features dialog. The Roman alphabet will be automatically entered into the Alphabetic Sequence dialog so you may close that dialog by clicking the OK button. Then in the main menu select Ontology and Lexicon, Lexicon, Nouns.

3.2 Entering Nouns into the Target Lexicon

We're now ready to begin entering some nouns into your target lexicon. First we'll enter several stems with their glosses. Then we'll define several features for the nouns and set their values appropriately for each particular noun. Finally we'll specify a form for the nouns and we'll write the necessary rules to generate that form.

3.2.1 Entering Noun Stems into the Target Lexicon

We'll begin by entering some noun stems into your target lexicon. To enter a stem into the lexicon, you must be in the Stems view, which is the view that comes up initially when you come to this screen. You'll see that the Stems button in the upper left corner is highlighted indicating that you're in the Stems view. To enter a new stem, click on the New Stem button in the upper center of the screen. Enter the stem *man* into the New Stem dialog and click OK. You'll then be asked to enter a gloss and comment. You should always enter a gloss for each stem. Since you're entering English stems for these tutorials, you can just enter the same word for each of the glosses. For these tutorials you can leave the comment entry blank. Enter each of the noun stems listed below. As you enter these stems, you'll see them displayed in the table in alphabetical order.

Noun Stems: woman, brother, baby, Mary, John, God, city, house, church, water, love, bush, dress, sister, Jerusalem, Egypt, king, hand, book, river, prayer

If you ever want to edit a stem, just right click on that stem. Similarly if you want to edit a particular gloss or comment, just right click on the appropriate cell. If you want to delete a stem, click on that stem and push the Delete key. You'll be asked to confirm the deletion.

3.2.2 Specifying Features for the Target Nouns

Now let's specify some features for these nouns. To enter features, you must be in the Features view. Click on the Features button in the upper left corner. You'll see that the Stems and Glosses columns remain in the table, but right now there aren't any features defined for nouns so no other columns will be displayed. First we'll enter a feature for gender so that later we'll be able to generate appropriate pronouns (e.g., he, she, it) for our nouns. Click on the New Feature button in the upper center of the screen, and enter Gender as the feature name. You'll then see a new dialog prompting you to enter values for this feature. The previous tutorial mentioned that all features consist of a name and a set of values. This dialog is where you can enter the values for your features. The first feature value that you enter will automatically be used as the default value for all the nouns you've entered into your lexicon. Since most nouns in English are neuter, enter **Neuter** into the first row. As soon as you push the Enter key, the system will suggest a single letter abbreviation for this feature value, in this case 'N'. You may change that single character abbreviation if you want, but you should generally accept the system's suggestion. Continue by entering **Masculine** and **Feminine** into the next two rows in the table. Then click the OK button. You'll now see a new column labeled Gender and each of the nouns will be set to Neuter because that is the default value. Go ahead and set the Gender value appropriately for each noun by clicking on the dropdown arrow and selecting the correct value (e.g., for *man* select Masculine). As you're setting the value for each noun, ask yourself, "If this noun were to be replaced by a pronoun, would the pronoun be *he*, *she* or *it*?" The word *baby* is gender neutral, so you should probably set its gender value to Masculine. Although *God* is neither masculine nor feminine, when *God* is represented with a pronoun, you probably don't want *it* to appear in the text, so set the gender appropriately.

After you've finished setting the Gender value for each of the nouns, add another feature called **Type**. This feature will have two values: **Common** and **Proper**. Enter these values into the feature value dialog and then set the value appropriately for each of your English nouns. This feature will eventually be used when we're writing rules to generate articles (e.g., *the*). Only the common nouns will be marked with the definite article; in English we never say something like *the Mary*¹.

These two features will be sufficient for our introductory English grammar. Take a moment and think about the features that you'll want to enter for your target nouns. If your language has noun classes, you'll want to enter a feature named Class and then enter the names for each class. You may want a feature to distinguish countable nouns from mass nouns. You may want another feature to indicate whether a noun is inalienably possessed or not. Some nouns may have an honorific status which needs to be signaled by a feature value. Before proceeding, think carefully about how you can use features to divide your target words into various categories that are relevant to your grammar.

If you ever want to change the name of a feature, right click on the column label for that feature. You'll then be prompted to edit the feature's name. If you want to delete a feature, click on any cell in that feature's column and push the Delete key. You'll be asked to confirm the deletion. If you want to add new feature values or change the feature values for a feature, click on the Edit a Feature button, select the feature that you want to edit in the dropdown at the top, and then edit the feature values. You may always add new feature values, and you may always edit feature value names. However, you should be careful about changing the single letter abbreviation for a feature value. You may change those if you like, but if you've written any rules that use those feature values, those rules will have to be updated by you.

3.2.3 Specifying a Form for the Target Nouns

Now let's enter a form for our nouns. To enter forms, you must be in the Forms view. Click on the Forms button in the upper left corner of the screen. Again you'll see the Stems and Glosses columns, but since there aren't any forms defined yet for nouns, no other columns will be displayed. To enter a new form, click on the New Form button in the upper center of the screen. Enter the name of the new form, in this case **Plurals**. You'll then see a column appear with that name. If you want to change the name, right click on the column header and you'll be prompted to edit the name. Now you'll need to write rules that will

¹ Actually we can say things like, "The Mary that I know would never do that." But these situations are rare and won't occur in the semantic representations.

generate this form. Click on the Form Rules button in the upper center of the screen. You'll see a dialog appear with the name of your form in the tree.

In English the plural form generally consists of the noun stem plus the plural suffix –s. To write a rule that will add this suffix, click on the Grammar Wizard button at the bottom of the Lexical Form Rules dialog. The Grammar Wizard dialog will appear. At the top of this dialog you'll see eleven rule templates as shown below in Figure 3-1.



Figure 3-1. Grammar Wizard dialog with Rule Templates

These eleven templates briefly summarize the types of rules you can write here in the lexicon. We want to write a rule that will add a suffix, so select the second template and double click. After double clicking on this entry, you'll see the dialog change as shown below in Figure 3-2.



Figure 3-2. Grammar Wizard Dialog building a rule to add a Suffix

You're now ready to enter the suffix, so enter **s** and push the Tab key. For the suffix's tag enter **Plural** and push the Tab key again. A nearly complete spellout rule will then be automatically built for you. To complete the rule, you need to specify a group in the dropdown at the upper right corner. Open that dropdown and there will only be one entry – Plurals. Select the Plurals group so that this

rule will generate the plural forms. We should also enter a comment to illustrate what this rule does. Even though this is a very simple rule, it's always a good idea to enter examples into the Comment box. Click in the Comment box and enter **Examples:**. When you're entering text into a rule's Comment or Name, you will undoubtedly want to enter some text in the target font and other text in the standard Arial font. When you first begin typing in either the Name or Comment box, the font will be Arial. To switch to the target font, push the Alt key. The text box's color will then change to light purple indicating that you'll be typing text in the target font. Begin typing and you'll see the text appear in the target font. So push the Alt key now and then type brother. The word will appear in the target font. Push the Alt key again and the text box will change back to white, indicating that you're typing in the standard Arial font. Type -> , and then push the Alt key again and type brothers. This text will appear in the target font. You may alternate back and forth between the standard font and the target font as many times as you like. Enter another example into the comment box such as river -> rivers.

At this point the rule is complete, but you are strongly encouraged to specify one or more grammar topics for each rule that you write. There are two benefits to assigning rules to grammar topics: 1) In a later lesson you will learn how to automatically search the grammar for all rules that are associated with a particular topic. This will be very helpful when your grammar has grown to the point where it has several hundred rules and you can't remember which rule performed a particular operation. 2) When you export your grammar to a text file, the grammar will be organized by topics. So by assigning each rule to one or more topics, your exported grammar will be more complete. To assign a rule to a grammar topic, click on the button labeled Topics at the bottom of the rule dialog. You'll see a suggested list of grammar topics. You may always add new topics to this list by clicking on the button in the lower left corner labeled Add Grammar Topic. A good topic name for this particular rule is Number which is already in the list, so you may select it. If you want to select additional topics for this rule, you may do that now or anytime in the future. After you've selected one or more grammar topics, click the OK button and you'll see the selected topics appear in the rule's Topic dropdown. Your rule is now complete and should appear as shown below in Figure 3-3.

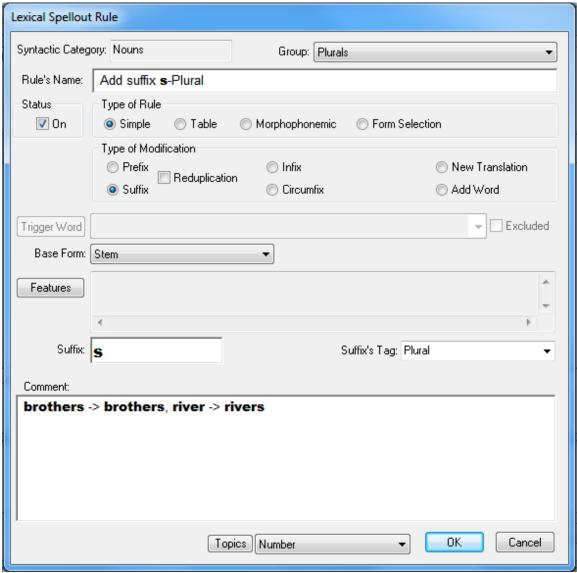


Figure 3-3. Lexical Spellout Rule that adds the suffix s

Let's take a minute to review what this rule does. This rule is in the Plurals group so it will generate noun forms that appear in the Plurals column. This rule adds the suffix –s to the stem. It's a very simple rule, but please take a moment to look at each of the controls and think about their contributions to the rule. This rule's Type is Simple meaning that it will add one affix, the Type of Modification is Suffix, the Base Form is Stem meaning that it will modify the stem of each word in order to build the plural form, and you've supplied a suffix and a tag for the suffix.

After you've entered this rule, click the OK button. You'll then see the results of your new rule in the Plurals column. If you check the Show Tags option in the upper left corner of the screen, you'll see the morpheme boundaries as well as the labels for the morphemes. You can always resize the width of any column by resting the cursor at the column label's right edge and then dragging

the column edge left or right. You can adjust the widths of form columns with the Show Tags option checked and unchecked.

Note: When entering rule names, rule comments, stem glosses, stem comments, and feature comments, you may enter text in the target font simply by pushing the Alt key. After you push the Alt key, begin typing and you'll see the text appear in the target font. If you want to switch back to the standard font, simply push the Alt key again. You may alternate back and forth between the standard and target fonts as many times as necessary just by pushing the Alt key. Whenever the text box's color is white, you'll be typing in the standard Arial font; whenever the text box's color is light purple, you'll be typing in the target font.

The rule that you just entered works fine for most nouns, but undoubtedly you've noticed some problems with several of the plural forms. We'll write a couple of morphophonemic rules to correct those problems, but first let's experiment with this rule just to see how it affects our plural forms. To edit the rule, click on the Form Rules button again, and then double click on the rule's name in the tree. The rule will then be displayed. Change the rule so that it adds a prefix s- rather than a suffix (note that you don't need to enter the hyphen). You'll need to enter a tag for the prefix so enter **Number** again. Save the rule by clicking OK and then check your plural forms again. Now they each have a prefix at the beginning. Just for fun, edit the rule again and this time check the Reduplication box that is between the Prefix and Suffix options. Then in the prefix dropdown, select 'Reduplicate first two characters'. Save the rule and look at your plural forms again. Now let's make the rule apply only to the masculine nouns. Edit the rule and click the Features button on the left side. You'll see a new dialog with the Gender feature that you defined earlier; select just the Masculine option. Save the rule and look at your plural forms. Now only the masculine nouns have the first two characters reduplicated. Edit the rule's features again, but this time select both the Masculine and Feminine options. After saving the rule and confirming that it's working the way you'd expect, edit the rule's features again. This time deselect the Masculine and Feminine options, and right click on the Neuter option. A right click indicates that you don't want that value included. You'll see the box around the check mark turn red indicating that that value won't participate in the rule. Save the rule and confirm that your plural forms are being generated as you've specified. When you specify features in a rule, you may select as many values as are relevant by left clicking, or you may block as many unwanted values as are irrelevant by right clicking. Go ahead and experiment with the rule any way that you want. When you're finished, put the rule back to its original form, but specify that it shouldn't apply to Proper nouns – we don't want to generate a plural form for names like Mary or John.

Now we need to write a couple of morphophonemic rules that will correct some of the plural forms. Let's write a rule that will correct the plural suffix for the word *church*. Click the Form Rules button and then the Grammar Wizard button.

Look at the eleven templates, in particular the fifth one because we want to write a morphophonemic rule. Select the fifth template and then double click it. This rule applies to a suffix so double click on **ends**. For the "End of Stem" enter **ch** and then push the Tab key. For the suffix enter **s** and push the Tab key again. For the suffix's tag double click on **Plurals**. We don't want this rule to change the stem, so enter **ch** again for the "New end of Stem" and push the Tab key. For the new suffix enter **es** and push the Tab key. In the rule's dialog select Plurals for the group and click the Topics button to select Number. In the Comment enter **Example: church+s** -> **church+es**. Your new morphophonemic rule should appear as shown below in Figure 3-4.

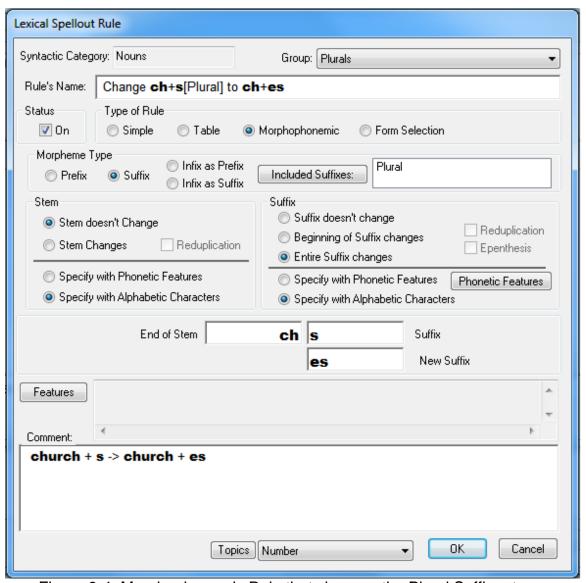


Figure 3-4. Morphophonemic Rule that changes the Plural Suffix s to es

Save this rule and confirm that the plural form of *church* is now being generated correctly. If we look at the plural form for *bush*, we see that the morphophonemic

rule we just entered can correct its plural form if we modify the rule slightly. Edit that morphophonemic rule by double clicking on its name in the tree, and then change the End of Stem entry to **ch,sh** (no space after the comma). Save the rule and confirm that the plural form of *bush* is now being generated correctly. To correct the plural form of *dress*, edit the rule again and change the End of Stem entry to **ch,sh,s**.

English has a spelling rule that changes plural nouns ending with a 'y' from 'y+s' to 'i+es' as in words like *families*. Let's use the Grammar Wizard to enter this rule. Go ahead and build that rule yourself; your completed rule should look like that shown below in Figure 3-5.

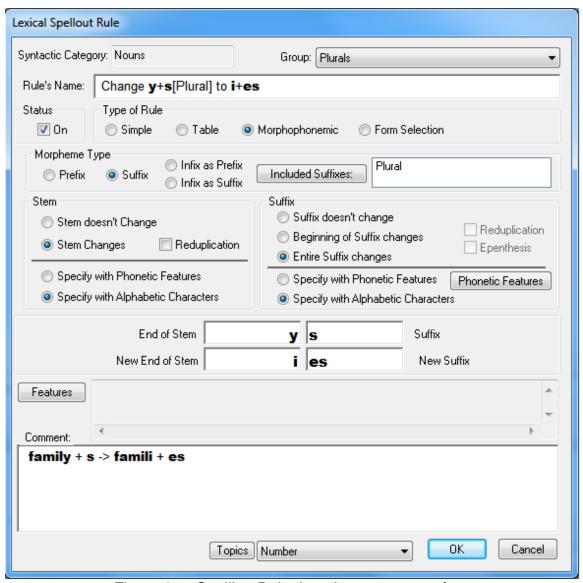


Figure 3-5. Spelling Rule that changes **y+s** to **i+es**

Now your list of plural nouns should be complete except for the suppletive forms. A suppletive form is any irregular word that cannot be produced by rules.

Most languages have suppletive forms — words which must be memorized because they cannot be generated by rules. When this is the case, simply right click on a cell that contains an incorrect form and begin typing the suppletive form into the resultant dialog. After you finish typing the irregular word, push the Enter key and you'll see the suppletive form highlighted with a white background. For example, the plural form of *man* is *men*, and the plural form of *woman* is *women* so you can enter those suppletive forms now. After you've entered a suppletive form, you may later want to erase it and let the program generate the form based on your rules. If this is the case, simply click on the cell and push the backspace key to delete the suppletive form, and then push the Enter key. You'll see the form generated by rules appear in the cell, and the cell's background color will return to light blue.

3.3 Entering Verbs into the Target Lexicon

Now let's switch to the verbs and continue this process. You can change to the verb syntactic category by selecting it from the dropdown in the upper left corner of the screen. Be sure to push the Stems button so that you can begin entering verb stems.

3.3.1 Entering Verb Stems into the Target Lexicon

Enter the English verbs listed below. In the list of verb stems, three include prepositions: *lift up*, *put away*, and *knock over*. When entering those verbs, enter the verb as the stem, and then click on the button labeled Add a Modifier. You'll then be prompted to enter the preposition as a modifier.

Verb Stems: walk, read, eat, see, give, say, go, ask, be, think, tell, make, lift up, put away, knock over, break, bake, choose, beg, cut, dig, punish, search, bury

When you enter a verb that includes a preposition, you'll need to supply a tag such as **Verbal Preposition**. The modifier dialog for *lift up* will look like that shown below in Figure 3-6 (remember that you will first enter *lift* as the verb stem and then click the Add a Modifier button).

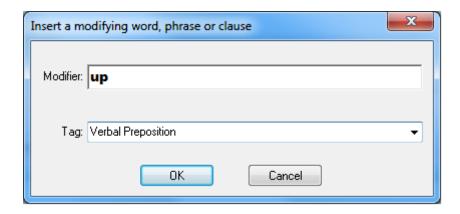


Figure 3-6. Modifier dialog when entering *lift up*

After you've entered a verb with an accompanying preposition, the program will check the first VP phrase structure rule to see if the modifier's position has been specified. Since this is the first Verbal Preposition, its position in the VP won't have been specified yet, so you'll be asked whether or not you want to edit the first VP phrase structure rule. When you're asked that question, answer Yes and you'll then see the phrase structure rule with Verbal Preposition entered at the top of the tree. If you click on Verbal Preposition in the tree, you'll see a popup showing an example of your Verbal Preposition. You may move the Verbal Preposition so that it follows the Verb simply by dragging it down and dropping it on the verb. We'll learn about the phrase structure rules in a subsequent tutorial, so we won't dwell on them now. Enter all of the verbs listed above into your target lexicon.

3.3.2 Specifying Features for the Target Verbs

After you've finished entering the verbs, add a feature for verbs called 'Complement Clause Structure'. Its values will be 1) Not Applicable, 2) Propositional, 3) 'to' Infinitive, 4) 'ing' Infinitive, and 5) Bare Stem Infinitive. We'll use this feature later when we start generating object complements. Enter these feature values just as you did the noun features, and specify the proper value for each verb (e.g., walk – Not Applicable, see – 'ing' Infinitive (Mary saw John reading that book.), beg – 'to' Infinitive (Mary begged John to read that book., etc.).

3.3.3 Specifying Forms for the Target Verbs

English verbs generally have four forms: Past Tense (e.g. *ate*), Past Participle (e.g., *eaten*), Gerund (e.g. *eating*), and Third Singular Present (e.g., *eats*). Add columns for these four forms by selecting the Forms view and then clicking the Add Form button. After you've added these four forms, you can use Grammar Wizard to enter a rule that will add the suffix *-ed* to the stem in order to generate the Past Tense form. Be sure to include an example in the comment, and assign the rule to one or more appropriate grammar topics. Your final rule should look like that shown in Figure 3-7.

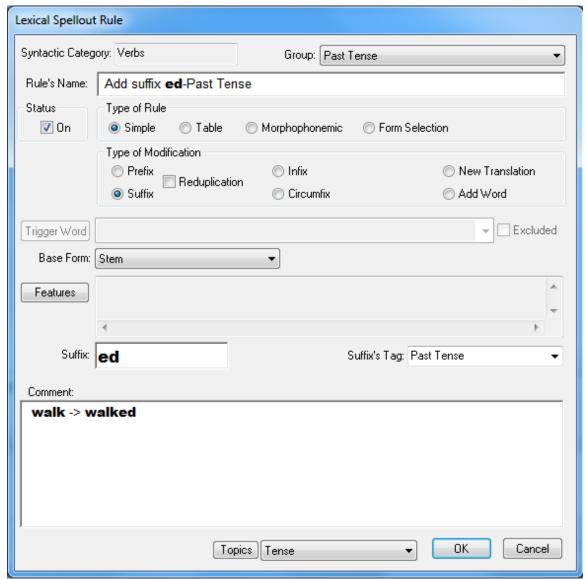


Figure 3-7. Lexical Spellout Rule that adds the Past Tense Suffix ed

After you've entered this rule, you'll need three morphophonemic rules; one will change e+ed to e+d (e.g., baked), and another will change y+ed to i+ed (e.g., buried). Go ahead and enter those two rules on your own. The third rule is a bit more complex; it needs to correct the form of beg. We need to write a morphophonemic rule that looks for stems that end with CVC and are then followed by ed. This rule is too complex for Grammar Wizard, so we need to write it ourselves. Begin by double clicking on Past Tense in the Lexical Form Rules dialog. You'll see an empty spellout rule dialog appear. Fill that dialog in as shown below in Figure 3-8. You can specify the phonetic features of the stem by clicking on a cell in the End of Stem grid. That grid represents the final five characters of the stem. We want the final character of the stem to be a consonant, so click on the last (right-most) cell in that grid. You'll then see a dialog with all of the phonetic features that were defined for us when we started

this language project. You can just select Consonant in the first column, and then click OK. You'll see a "C" appear in the cell that represents the last character of the stem. The character that precedes the last character must be a vowel, so click in that cell and select Vowel in the first column of the Phonetic Features dialog. The character before that vowel must be a consonant, so repeat the process for specifying a consonant. The final result should look like the rule shown below in Figure 3-8.

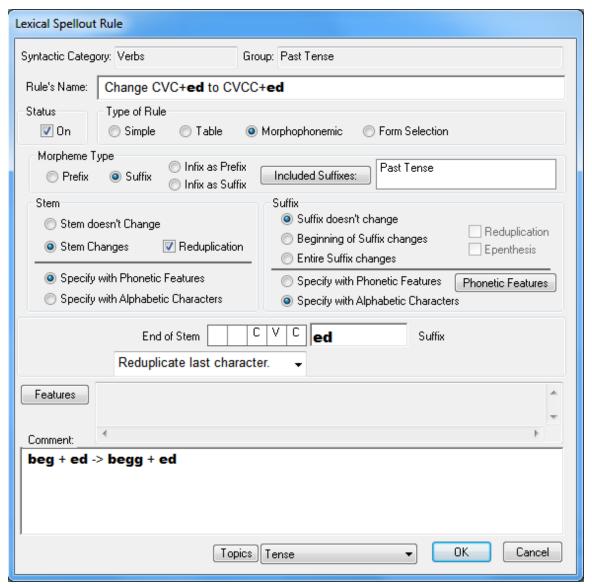


Figure 3-8. Morphophonemic Rule that changes CVC+ed to CVCC+ed

The rule shown above will work for many verbs (e.g., begged, committed, controlled, dipped, hugged, stopped, tripped, etc.), but there are exceptions to it (e.g., delivered, entered, gathered, etc.). All of the exceptions will be considered suppletive. To enter a suppletive form (e.g., broke, bought, chose, cut, dug, etc.), right click on the appropriate cell, and then enter the suppletive form into

the resulting dialog. Don't be concerned about entering the past tense form for be. Later we'll write a table spellout rule specifically for be. All of your past tense forms should now appear correctly in your lexicon.

For most English verbs the Past Participle form is identical to the Past Tense form (e.g., *John has asked/begged/buried/*etc. ...). However, there are many verbs that have a Past Participle form that is different than the Past Tense form (e.g., *John has eaten/broken/seen/*etc. ...), so we need to include this form in the lexicon. To generate the Past Participle form for each verb, we need a rule that will copy the Past Tense form into the Past Participle column. If you look at the Grammar Wizard templates, you'll see that the fourth template accomplishes this task; it builds a form selection rule. Select the fourth template and then double click on it in the prompting box. Then double click on "Past Tense form" in the prompting box. You'll then see a slightly different type of lexical spellout rule. This is a form selection rule; all it does is copy a previously defined form into the current form. Assign this rule to the appropriate group "Past Participle," assign it to one or more relevant grammar topics such as "Tense," and then enter an example. Your final rule should look something like that shown in Figure 3-9.

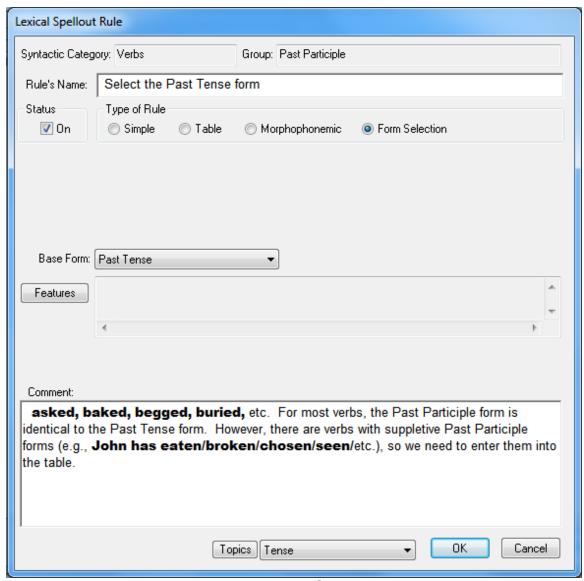


Figure 3-9. A Form Selection Rule

After you've entered the rule shown above for the Past Participle form, you'll need to type in the irregular forms for many of the verbs (e.g., *eaten, seen,* etc.). Go ahead and do that now.

To generate the gerunds, use Grammar Wizard to write a rule that will add the suffix -ing to the stem. Be sure to put that rule in the Gerund group. Then write a morphophonemic rule to correct the form for verbs like choose, give, bake, etc. That rule will delete the final e before -ing. You'll need to write one more morphophonemic rule to correct the form for verbs such as begging, cutting, digging, etc. That rule will reduplicate the final consonant for verbs that end with CVC. That rule is very similar to the rule that we wrote for Past Tense, so it won't be repeated here. After writing this morphophonemic rule, the gerund forms of begging, cutting, digging, etc., will be corrected. However, look at the gerund forms of bake, give, and make. They had their final consonants

reduplicated because the initial morphophonemic rule deleted the stem final e. The best solution to this problem is to order these two morphophonemic rules so that the stem final consonant reduplication rule applies first, and then the stem final e deletion rule applies second. You'll find that rule ordering is often crucial when writing a grammar for a language, so reordering the rules is very easy in TBTA. To reorder these two rules, click on the Form Rules button, and then in the Lexical Form Rules dialog, click on the rule that deletes the stem final e, and drag it downward and drop it on the rule that reduplicates the final consonant. Whenever you need to reorder rules, you can simply drag and drop them to the desired location. Close the Lexical Form Rules dialog, and you'll see that the gerund forms are now correct.

For the third singular present form, write a rule that will add the suffix –s to the stem. The tag for this suffix should be something like Third Singular Present. Then write a morphophonemic rule for verbs like *punish* and *search*. The End of Stem entry for that morphophonemic rule will look like: **sh,ch**. Then write another morphophonemic rule to account for verbs that end with *y* such as *bury*. You'll then need to correct the third singular form of *say*; we'll treat that as suppletive for now. And the third singular form of *be* is also suppletive so enter it into the table by right clicking on the cell. Now all of your third singular present forms should be correct.

3.4 Conclusion

You've now entered all of the features, forms and form rules that are necessary for English nouns and verbs. Spend some time thinking about the features and forms required for your target language.

For the sake of completeness, go ahead and enter a few more stems into some of the other syntactic categories. For these tutorials, we won't specify any features or forms for these other categories.

Adjective Stems: big, good, red, three, third

Adverb Stems: carefully, quickly Conjunction Stems: and, or, but

In this lesson you've learned how to enter target stems into the lexicon. You've also specified features and forms for those stems, and you've written Lexical Spellout rules to generate those forms. Whenever a particular stem has a suppletive form, you've entered that form directly into the lexicon. These rules that you've entered are called Lexical Spellout rules because they're in the lexicon and they spell out, or generate, the various forms. In the next lesson you'll learn about the spellout rules in the grammar. Those rules are virtually identical to these rules, but because they're in the grammar rather than the lexicon, they serve a slightly different purpose. That purpose will become clear in the next tutorial.