

TBTA Tutorial Lesson 9: Complex Concept Insertion Rules and Feature Adjustment Rules

(video: [TBTA-9A Complex Concept Insertion Rules](#))

(video: [TBTA-9B Feature Adjustment Rules](#))

9.0 Introduction

In the previous lessons you've learned about the ontology, the semantic representations, the target lexicon, and all of the grammatical rules in TBTA's synthesizing grammar. In this lesson we're going to begin looking at the final major component in TBTA – the transfer grammar. The purpose of the transfer grammar is to restructure the semantic representations in any way that's necessary in order to build a new underlying representation that's appropriate for your target language. The semantic representations contain concepts that have been lexicalized by English, and they consist of structures that resemble English sentence structures. But other languages employ different structures, and their lexical items reflect a different view of conceptual space. So the transfer grammar must build a new underlying representation that consists of the target language's words, structures, and features. The transfer grammar contains nine different types of rules; this lesson will introduce two of them: complex concept insertion rules and feature adjustment rules.

Linguists are very familiar with synthesizing grammars. In your linguistics classes you've probably seen rules similar to the ones discussed in the previous tutorials. In your other classes you've probably discussed phrase structure rules, agreement rules, spellout rules, morphophonemic rules, etc. However, linguists don't discuss transfer grammars because they're generally concerned with language description rather than translation. Since TBTA is ultimately a translation project, we must include a transfer grammar. Since you have not talked in your other classes about the rules that you'll see here in the transfer grammar, these tutorials will walk through them and illustrate them very clearly.

9.1 Complex Concept Insertion Rules

The tutorial that discussed the ontology described the concepts that are permitted in TBTA's semantic representations. Generally only semantically simple concepts are used in the semantic representations in order to reduce the problem of lexical mismatch between the source and target languages. However, if only semantically simple concepts are used, the texts generated by TBTA become unnecessarily long and don't communicate the message faithfully. For example, the concept SHEPHERD is semantically complex, and many languages don't have a lexical equivalent. So whenever the word *shepherd* occurs in a source text, it is explicated as *man who takes care of sheep*. However, speakers of languages which have a lexical equivalent for SHEPHERD don't want to read texts that contain *man who takes care of sheep*, they want the texts to talk about *shepherds*. Therefore complex concepts are explicated

when building the semantic representations, but complex concept insertion rules are able to insert semantically complex concepts back into the semantic representations if the target language has a lexical equivalent.

Go to Luke 2:8 and look at the semantic representation. You'll see that it produces *That night some men who took care of sheep were in a field that was near Bethlehem. Those men were protecting their sheep.* In order to activate the complex concept insertion rule for SHEPHERD, click on the button labeled Complex Concepts in the lower right corner of the lexicon and grammar development screen. You'll then see the dialog that is shown below in Figure 9-1. That dialog lists all of the complex concepts that have insertion rules.

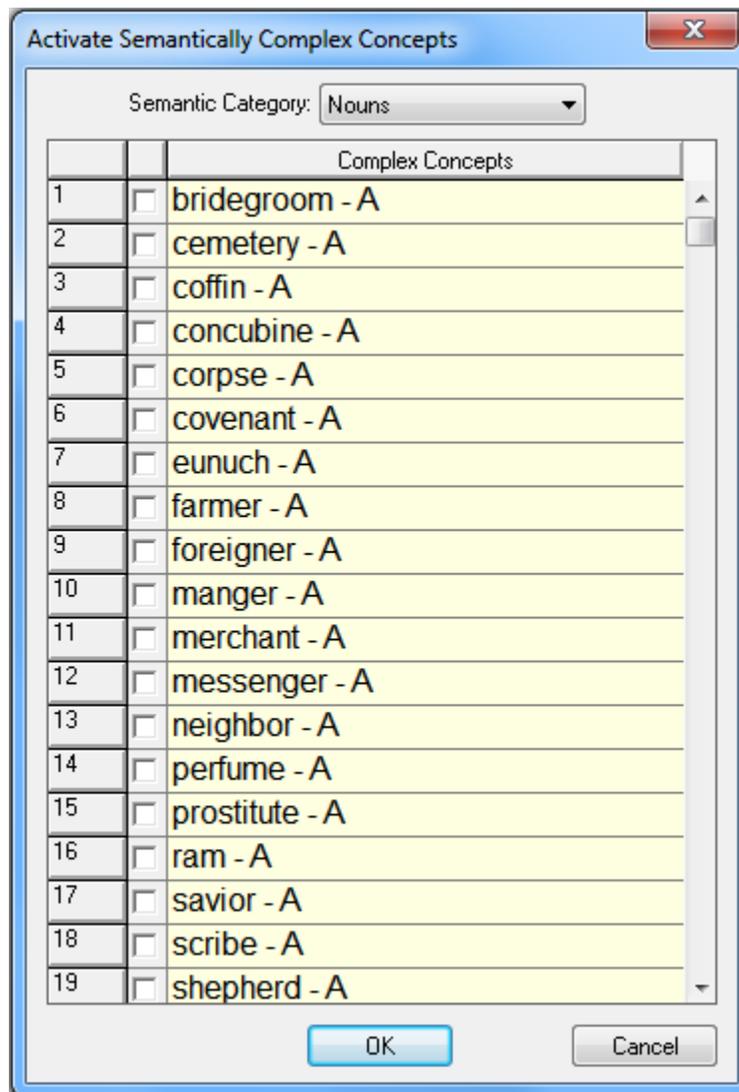


Figure 9-1. Complex Concept Insertion Dialog

There's a dropdown at the top of this dialog so that you can see the complex concepts in other syntactic categories. If you open that dropdown, you'll see that there are semantically complex nouns, verbs, and adjectives. Select Nouns and you'll see that one of the complex concepts is SHEPHERD-A. Check the box next to SHEPHERD-A,

and then click the OK button. You'll then see that the semantic representation in Luke 2:8 is changed to produce *That night some shepherds were in a field that was near Bethlehem. Those shepherds were protecting their sheep.* Now click the button labeled Complex Concepts again, and uncheck the box next to SHEPHERD-A. Close the dialog and you'll see that the semantic representation returns to its original form.

For an example of a semantically complex adjective, go to Luke 5:18. Look at the semantic representation and you'll see that the first sentence produces *Some men decided to bring a man that was not able to move his legs to Jesus.* English has the word *paralyzed*, so English speakers don't want to read texts containing relative clauses like the one in this verse. Open the Complex Concepts dialog again, select Adjectives in the dropdown at the top, and activate the rule for PARALYZED-A. Then close the dialog and look at the semantic representation again. Now the first sentence produces *Some men decided to bring a paralyzed man to Jesus.* This is a better translation for languages that contain a lexical equivalent for *paralyzed*.

Whenever you begin a new language project, you should open this dialog and go through the lists of complex concepts and decide which ones have a good lexical equivalent in your target language. For each complex concept that has a good lexical equivalent, activate the associated complex concept insertion rule so that the generated texts will contain those complex concepts.

9.2 Feature Adjustment Rules

The tutorial that discussed the features in the semantic representations listed the features and all their possible values. However, no language will need all of those features nor all their values. Feature adjustment rules have been included in TBTA's transfer grammar so that you can tailor the features in the semantic representations appropriately for your target language. Specifically these rules permit you to 1) hide features that are irrelevant to your target language, 2) collapse feature values that are unnecessary in your target language, and 3) generate new features that are helpful in your target language. These three tasks are all performed with a single dialog which you can access by going to the Lexicon and Grammar Development screen, and double clicking on the 'Feature Adjustment Rules' node in the grammar tree. You'll then see the dialog that is shown below in Figure 9-2.

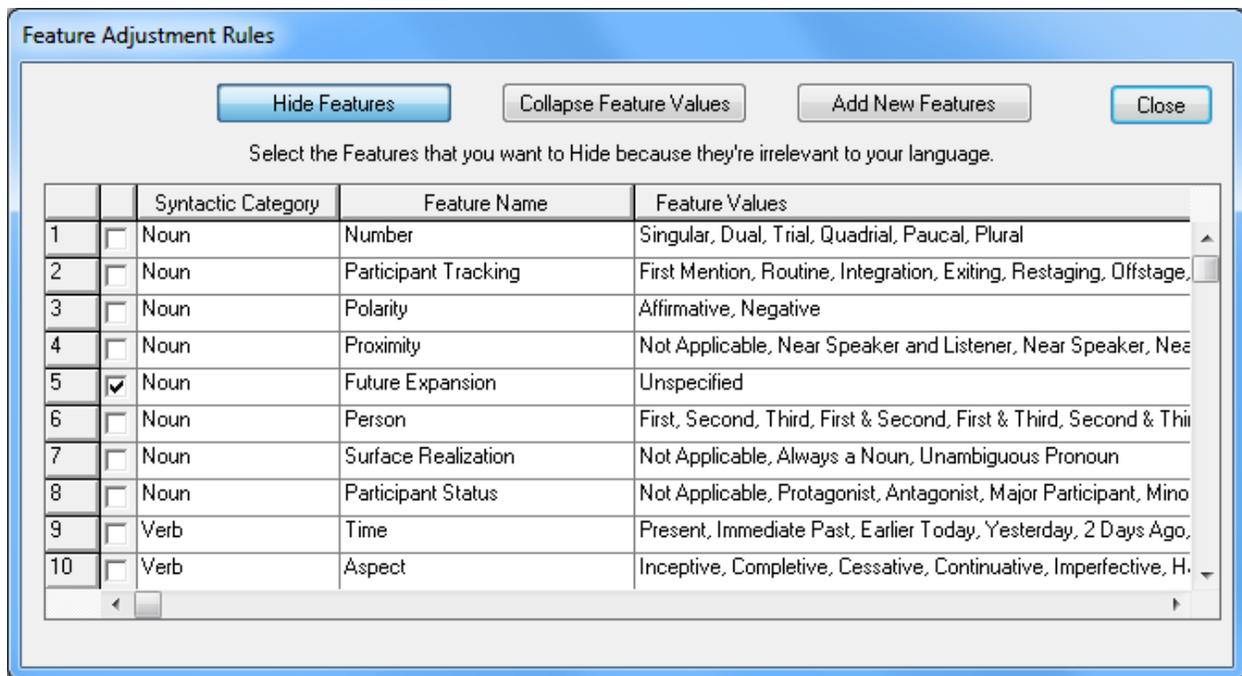


Figure 9-2. Feature Adjustment Rules Dialog

This dialog has three purposes which you can see in the three buttons at the top: 1) hide irrelevant features, 2) collapse unnecessary feature values, and 3) generate new helpful features for your target grammar. When this dialog is initially displayed, it has the Hide Features button highlighted indicating that you can specify features that should be hidden.

9.2.1 Hiding Irrelevant Features

The tutorial that described the features in the semantic representations listed five features that are pertinent to direct speech: 1) Speaker, 2) Listener, 3) Speaker's Attitude, 4) Speaker's Age, and 5) Speaker-Listener Age. When a language has multiple speech styles, these features may be used to determine which speech style is appropriate for a particular situation. However, not all languages have multiple speech styles. If your target language doesn't need to access these features, you can hide them so that they won't appear in either the semantic representations or in your grammar.

To hide features, go to the dialog shown above in Figure 9-2, and make sure the Hide Features button is highlighted. This dialog lists all of the features associated with each syntactic category. Scroll downward and you'll see that some features were selected to be hidden when you created your target language database. For example, all of the clause features associated with direct speech are hidden by default. If your language needs these features, you should uncheck the boxes in this table.

Scroll the table down to the section of features that are associated with clauses. Then, while this dialog is still open, rest your cursor on an opening clause boundary in the semantic representation. You'll see a popup explaining the features' values for that particular clause. Now hide one of the clause features by clicking the appropriate box in

this table; for example, click the box in the last row which is associated with Alternative Analysis. Now rest the cursor on the opening clause boundary in the semantic representation again, and you'll see that Alternative Analysis is not included in the popup. Now uncheck the box associated with Alternative Analysis, and rest your cursor on the opening clause boundary again, and you'll see that Alternative Analysis is included in the popup. Use this dialog to hide all of the features that are irrelevant to your target language.

9.2.2 Collapsing Unnecessary Feature Values

No language will use all of the feature values that are included in the semantic representations. For example, in the semantic representations nouns are marked for Number, and the possible values are Singular, Dual, Trial, Quadrial, Plural, and Paucal. However, very few languages will use all of those values; your target language may only morphologically distinguish singular and plural. If that's the case, the synthesizing grammar for your language should only have to deal with Singular and Plural; it should not have to deal with the values Dual, Trial and Quadrial. In order to collapse the values Dual, Trial, and Quadrial with Plural, click the button labeled 'Collapse Feature Values' at the top of the Feature Adjustment dialog. You'll then see a new panel with a syntactic category dropdown on the left, and a feature dropdown in the middle. The Number feature for nouns is the first feature for nouns, so it will be displayed when you click the 'Collapse Feature Values' button. In the tree below the two dropdowns you'll see all of the values for Number. If you want Dual to be treated as Plural, select Dual and drag it down to Plural, and then drop it. Your dialog will then appear as seen in Figure 9-3 below.

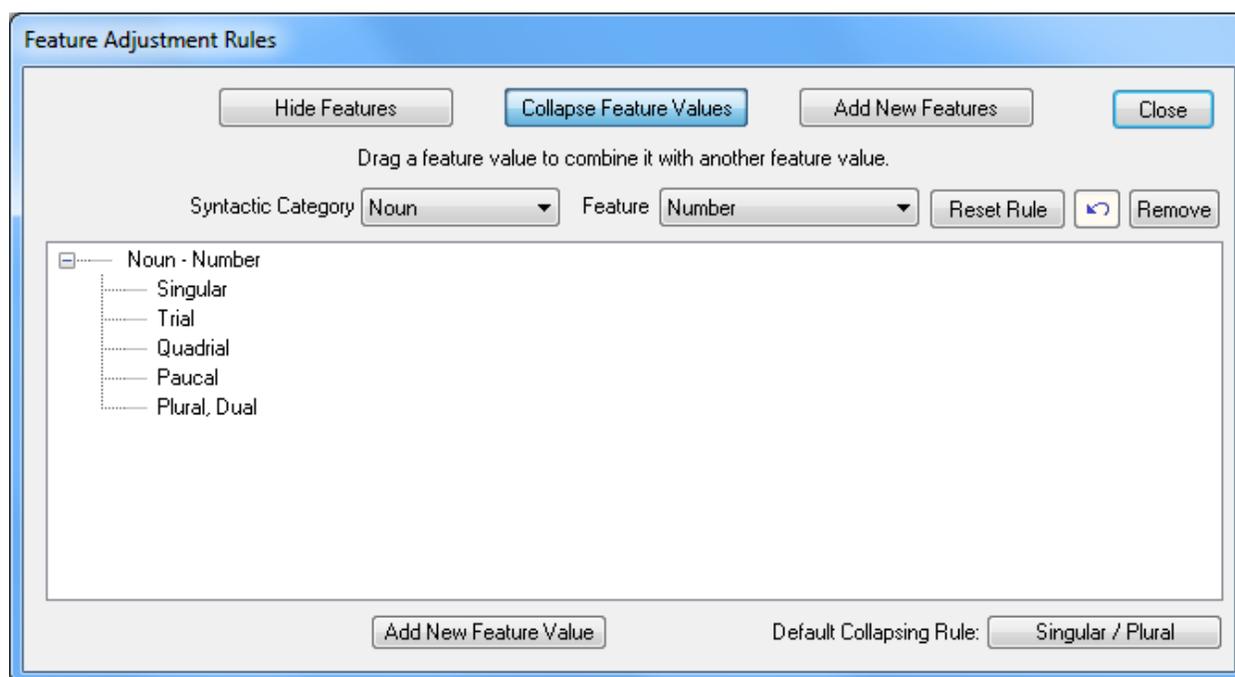


Figure 9-3. Feature Collapsing Dialog

Repeat this process and combine Trial and Quadrial with Plural. Then click the Close button in the upper right corner of this dialog, and go to Adjectives 1:5. That verse produces *John read three books*. Rest the cursor on the features below BOOK-A and you'll see that the popup says that the Number value is Plural. Then rest your cursor on the target word below BOOK-A, and the popup will tell you that the Number value was changed from Trial to Plural by the feature collapsing rule. Now open one of the spellout rules for nouns that you wrote in a previous tutorial. Click the Features button on that rule's dialog, and look at the values in the Number column. You'll see that the values Dual, Trial, and Quadrial are no longer there. You'll no longer see these values anywhere in the semantic representations or grammar rules. If you later decide that you need these values of Number, you can return to the dialog in Figure 9-3, select the feature in the dropdowns, and then click the Reset Rule button in the upper right corner. For example, close the spellout rule, and open the Feature Adjustment dialog again. Then click the 'Collapse Feature Values' button at the top. You'll see the Number feature for nouns again, and the values Dual, Trial, and Quadrial are beside Plural. Click the 'Reset Rule' button. You'll then see all the Number feature values restored. If you open a spellout rule for nouns again and click the Features button, this time in the Number column you'll see all of the original values.

In the 'Syntactic Category' dropdown on the Feature Collapsing dialog, select Verb. The first feature is Time, and it also has more values than any language will use. For this feature you'll probably want to define some values that are pertinent to your language. For example, you may want to define values such as "Past," "Remote Past," "Future," "Immediate Future," etc. To add your own values to this list, click the button at the bottom labeled 'Add New Feature Value'. Then enter your own value such as 'Remote Past.' Your value will be added to the list at the bottom. If your language considers anything that happened more than a year ago to be remote past, then drag and drop those values to 'Remote Past.' If your language only uses Past, Present, and Future, then you can use the shortcut button in the lower right corner labeled "Past / Present / Future." If you click that button, it will add two new values, "Past" and "Future," and then collapse the other Time values appropriately.

At this dialog you must decide what to do with the value of Time called Discourse. English speakers tell their stories using past tense, but that isn't universal. In the semantic representations, the vast majority of the verbs have a Time value of Discourse; only the verbs in direct quotes have actual time values¹ (e.g., two days ago, one week from now, etc.). Drag the value called Discourse to the appropriate value for your target language.

9.2.3 Adding New Features

As you've seen several times in these tutorials, the NPs in the semantic representations are marked with semantic roles rather than grammatical relations. However, the grammars for many languages refer to grammatical relations rather than semantic roles (e.g., English verbs agree in person and number with the subject, not the

¹ Actually, the verbs in relative clauses and adverbial clauses also have absolute time values, but this issue will be discussed in a subsequent tutorial.

agent). The third button at the top of the dialog in Figure 9-2 is labeled ‘Add New Features.’ Click that button and you’ll see the panel that is shown below in Figure 9-4.

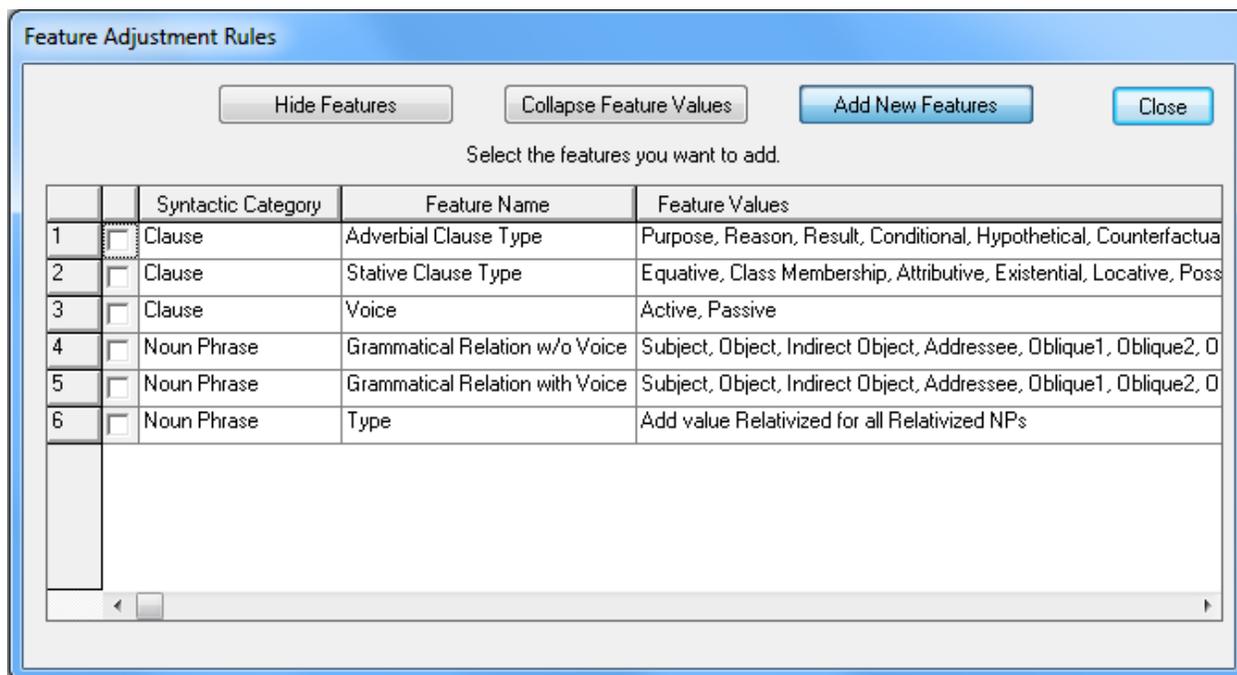


Figure 9-4. Add New Features Dialog

This dialog has five² predefined features that are very helpful when developing grammars. You will almost certainly want to activate rows 1, 2, and 6. Row 1 will add a feature to all clauses called ‘Adverbial Clause Type.’ The value of this feature will be ‘Not an Adverbial Clause’ for clauses which aren’t adverbial, but for clauses which are adverbial, this feature will specify the adverbial type (e.g., conditional, hypothetical, counterfactual, purpose, reason, etc.). This feature is very helpful because many adverbial clause types require particular verb forms. Row 2 adds a feature called ‘Stative Clause Type’ to all propositions. The value of this feature will be ‘Event (non-stative)’ for all events, but for stative propositions, this feature will specify the type of state (e.g., existential, locative, possessive, temporal, etc.). Row 6 will add a feature to all NPs indicating whether or not it contains a noun in a relative clause that is coreferential with the head noun. This feature is very helpful when dealing with relative clauses.

Row 3 adds a feature called Voice, and it is useful if your language has a grammatical method of promoting the patient NP to subject (i.e., distinguishing active from passive). If your language has a mechanism for promoting patients to subject position, you should activate rows 3 and 5. Rows 4 and 5 add a feature called ‘Grammatical Relation’ to all NPs, but row 4 ignores the proposition’s Topic NP value when assigning grammatical relations, while row 5 includes the proposition’s Topic NP when determining which grammatical relation to assign each NP. In general when you

² There are six rows in this table, but rows 4 and 5 both add grammatical relations. Row 4 is used for languages that don’t have any method of promoting patient NPs to subject, and row 5 is used for languages that do have a method for promoting patients to subjects.

start a new language project, you should activate rows 1, 2, 3, 5, and 6 unless you have a good reason not to.

9.3 Conclusions

This tutorial has introduced you to two types of rules in TBTA's transfer grammar: complex concept insertion rules and feature adjustment rules. When you start a new language project, you should walk through both of these types of rules early in the process. You can always come back later and change your values in these two types of rules, but these rules significantly impact your grammars, so you should deal with them early.