TBTA Tutorial
Lesson 1: Introduction

1.0 Introduction

Welcome to The Bible Translator's Assistant (TBTA). TBTA is a natural language generator designed specifically for field linguists who do translation work in minority languages. A natural language generator is a computer program that takes a semantic representation of a text, and a lexicon and grammar from a language, and then generates text in that language that has the same meaning as the semantic representation. If you learn how to use this generation system well, it will save you many years of work. TBTA will guide you as you develop a lexicon and grammar for your language. After your lexicon and grammar are sufficiently developed, TBTA will generate drafts of all the analyzed source materials. Then you and your mother tongue assistants will be able to edit those generated drafts into publishable texts.

There are three fundamental components in this generation system: 1) the semantic representations, 2) the target lexicon, and 3) the target grammar. The semantic representations will be provided for you, but you need to become familiar with their format. The format of these semantic representations will be the main topic of this introductory lesson. TBTA will walk you through the process of developing your target lexicon and grammar. The development of the target lexicon and grammar will be the focus of the remaining nine lessons. After you've completed these ten lessons, you'll have a good understanding of how the system works. However, it's not possible to include all of the system's details in these introductory tutorials; you'll need to learn those details on your own. Those details are provided in the context sensitive help system included with TBTA. In particular you'll want to consult the help system in order to learn more about the grammar rules. This help system will be described in subsequent lessons.

2.0 Format of the Semantic Representations

A semantic representation is any formal method of representing meaning. Linguists have proposed many different formats for semantic representations. However, because TBTA deals with minority languages, a new format had to be developed that is specifically oriented toward those languages. These semantic representations have been developed using a controlled English based metalanguage[1] augmented by a feature system designed to accommodate a very wide variety of languages. These semantic representations are comprised of 1) concepts, 2) features and 3) structures. For example, the semantic representation for the proposition *I should finish reading these books* is shown below in figure 1.

[1] A metalanguage is any language that is used to describe or analyze another language. In a bilingual dictionary where Korean words are described in English, English is considered a metalanguage because it is being used to describe Korean.

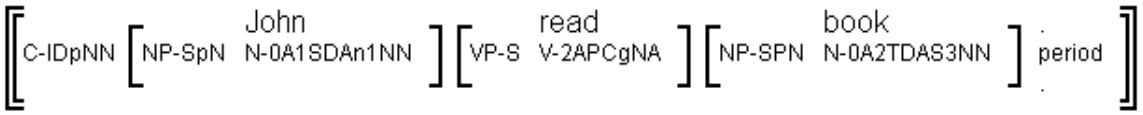

Figure 1. Semantic Representation of *I should finish reading these books*.

As seen in figure 1, there are three concepts in this proposition: JOHN, READ and BOOK. In these tutorials concepts will always be written in capital letters to distinguish them from English words. The features in the semantic representation are indicated by the sequences of characters below each concept and next to each proposition and phrase boundary. The structure is indicated by the proposition boundaries and phrase boundaries. You can see the semantic representation of a proposition by selecting in TBTA's main menu Generator, Lexicon and Grammar Development. At that screen you'll see a large window that contains the semantic representation for the currently selected verse. The rest of this tutorial will explain the concepts, features and structures that are used to build these semantic representations.

2.1 Overview of the Concepts used in the Semantic Representations

A concept is a structured idea. For example, English speakers are familiar with the concept of "a written message that is sent or delivered to someone." This particular concept has been lexicalized in English with the word *letter*. For convenience TBTA's semantic representations use concepts that have been lexicalized by English, and English labels are used to represent those concepts. Therefore this particular concept will be represented by LETTER. But the English word *letter* has other senses or meanings. For example, the word *letter* can refer to a character in an alphabet. In order to distinguish the written message from the character in an alphabet, the written message will be represented by LETTER-A and the alphabetic character will be represented by LETTER-B. So *letter* is an English word and it has at least two senses or meanings. LETTER-A has just one meaning; LETTER-A represents the concept of "a written message that is sent to someone." Because TBTA's ontology precisely identifies the meaning of each concept, you'll be able to easily map these concepts to target words or constructions.

When discussing objects such as *letters*, the various senses are generally quite clear. When discussing events, attributes or relations, the various senses can be more difficult to distinguish. For example, consider the English word *open*. This word has many different senses, two of which are: 1) to move a door, window or gate of some type in a certain way so that people or things may move in or out of a room, building, city, etc., 2) to remove the lid or cover of a container so that things may be put in or taken out of the container, etc. Most languages will probably require two different verbs for these two senses of *open*. Therefore the ontology has OPEN-A and OPEN-B. However, in English we can also say things like *John opened his eyes*, *John opened his mouth*, or *John opened a book*. These senses of *open* are quite distinct from OPEN-A and OPEN-B so your target language may require different verbs for these senses of *open*.

However, it's impractical to define an additional sense for each object that is being opened. Therefore concepts have been chosen which represent the major senses of each event, attribute and relation in the ontology. Throughout the semantic representations OPEN-A has been used when a person opens his eyes or mouth. If your language requires different target verbs for those senses of *open*, you'll need to write one or more collocational correction rules. Those rules will be described in a subsequent tutorial.

Ontology is the philosophical study of the nature of existence. When used in the context of natural language generators, the word 'ontology' refers to a structured list of concepts. To see the ontology in TBTA, select Lexicon in the main menu, and then select Source to Target Mappings, Objects to Nouns. You'll see a screen with two grids. The upper grid contains all of the concepts in the ontology; the lower grid will eventually contain your target lexicon. A small section of the ontology is shown below in figure 2.

	Concept Stems	Senses	Mappings	English Glosses
554	leprosy	A		the skin disease
555	letter	A		a letter, note
556	letter	B		a letter of the alphabet
557	Levi	A		Jacob's son

Figure 2. The Ontology

The ontology is organized into seven semantic categories: 1) objects, which are realized by English nouns, 2) events, which are realized with English verbs, 3) object attributes which are realized by adjectives, 4) event attributes which are realized by adverbs, 5) relations which are realized by English prepositions, 6) conjunctions and 7) particles. In order to keep these tutorials simple, these semantic categories will be called by their standard syntactic equivalents: nouns, verbs, adjectives, adverbs, adpositions, conjunctions and particles. Similarly object phrases will be called noun phrases, propositions will be called clauses, etc.

As you work with the concepts in the ontology, be sure to carefully read the gloss provided for each concept. The concepts are always represented with English words, but the actual concept may be somewhat different than your idea of what that word generally means. For example, open the dropdown at the top of the screen and select Relations. The first four relations don't have lexicalized English equivalents so they're given descriptive names. An example of each of these relations follows:

1. ITERATION

[Clause [NP JOHN] [VP HIT] [NP PETER] [NP ITERATION [AdjP 3] TIMES]]

John hit Peter three times.

2. NAME

[NP FRIEND [NP OF JOHN] [NP NAME MARK]] …

John's friend named Mark …

3. QUANTITY

[Clause [NP JOHN] [VP DRINK] [NP WATER [NP QUANTITY GLASS]]]

John drank a glass of water.

4. TITLE

[NP ISAIAH [NP TITLE PROPHET]] …
Isaiah, the prophet, …
As indicated by the name, relations signal some type of semantic relationship between two objects or events. Scroll downward until you come to the relation WHEN-B. Its gloss is 'when-because + event specifies the time and reason for the main event'. An example of this relation occurs in the sentence *When the sink overflowed, I called a plumber*. The *when* in that sentence signals both a temporal relationship between the two events as well as a reason relationship. This is a very different sense of *when* than in the sentence *John was six when he learned to read*. The *when* in that sentence signals just a temporal relationship. Other languages will probably treat these two senses of *when* differently. The important point here is that you should look at the gloss of each concept to make sure that you understand it before you deal with it in your target language. The ontology includes twenty-one senses of OF and twenty-three senses of BE. No language will have twenty-one different ways of signaling possession or twenty-three different types of stative clauses, but you should make sure that you understand each individual concept before you enter a target word or rule that will be triggered by that concept.

Go back to the Event category in the ontology and scroll the upper grid downward. You'll see that the Lexical Sense column has three different colors. An example of these colors is shown below in figure 3.

	Concept Stems	Senses	Mappings	Case Frames	English Glosses
242	destroy	A		AB___f___	to destroy something, annihilate (AP)
243	detest	A		AB_____	to detest something (AP)
244	develop	A			
245	die	A		A_____	to stop being alive (p)
246	dig	A		Ab____g__	to dig in/on/from something (don't include 'hole') (AP)
247	dip	A		AB__E____	to dip something into a liquid (APD)

Figure 3. Three Categories of Events in the Ontology

The events in the ontology have been divided into three categories. The first category is called Semantic Primitives. Cognitive semanticists have developed a theory called Natural Semantic Metalanguage theory. They have identified a small set of concepts which they claim are innate and therefore universal. They believe that every language has a lexical or morphological equivalent for each of the semantic primitives. The semantic primitives are identified in TBTA with a purple background in the Senses cell; see DIE-A in figure 3 above. The next category is called Semantic Molecules and they're identified with a light yellow background in the Senses cell. DESTROY-A, DEVELOP-A and DIG-A are all semantic molecules as seen in figure 3. Semantic molecules are non-primitives which occur frequently in the definitions of other words. These molecules are semantically more complex than the primitives, but yet they're still considered semantically simple. Other languages are likely to have good lexical equivalents for the semantic molecules. The final category is called Complex Concepts. These concepts are identified by the dark yellow background; see DETEST-A and DIP-A in figure 3 above. The complex concepts are less likely to have good lexical equivalents in other languages than are the semantic molecules.

Therefore you'll need to deal with the complex concepts in a special way. This will be discussed thoroughly in the final tutorial.

2.2 Overview of the Feature System used in the Semantic Representations

In TBTA each semantic/syntactic category has a set of features that are specifically relevant to it. Each of these features has two parts: 1) a name and 2) a set of possible values. For example, one of the features associated with nouns is named Number, and the possible values are Singular, Dual, Trial, Quadrial and Plural. As you develop your lexicon and grammar, you'll define additional features that are relevant to your particular language. The features that you define will be assigned to specific syntactic categories, and they will consist of a name and a set of possible values. Features are displayed in the semantic representations as you saw above in figure 1. When you rest your cursor on the features, you'll see a popup explaining what each of those characters means. Shown below in figure 4 is the same semantic representation that was shown above in figure 1, but this time the popups explaining the features have also been included.

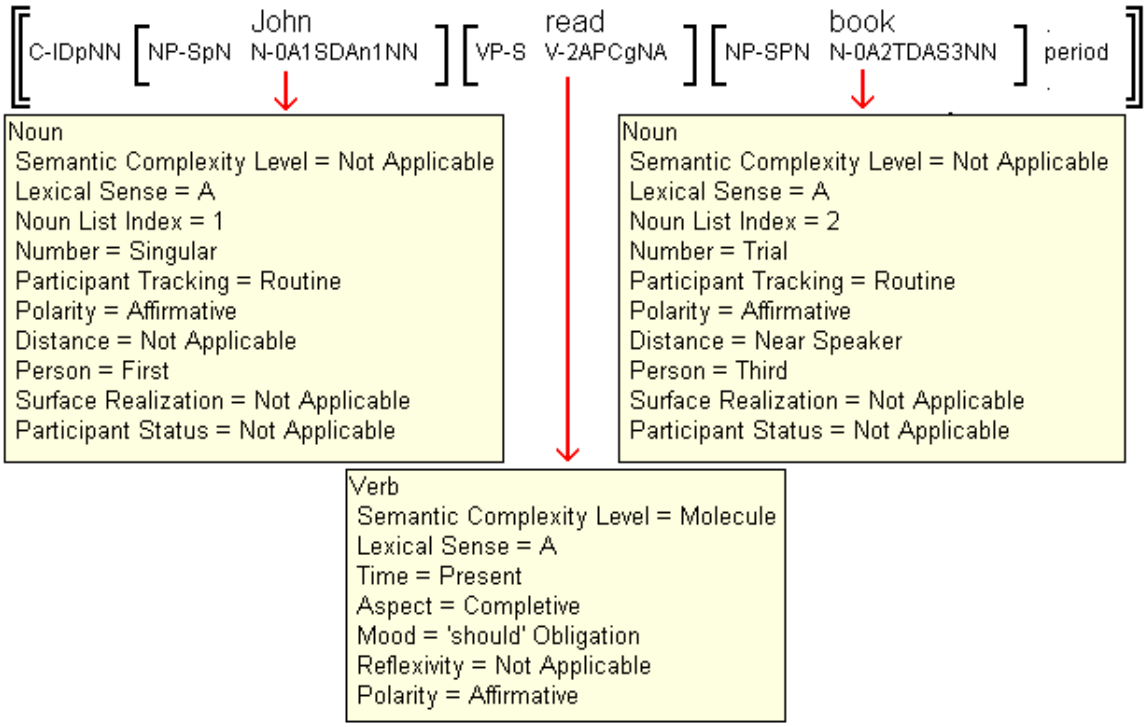

Figure 4. Semantic Representation with Popups Explaining the Features

For each concept the first two features will always be Semantic Complexity Level and Lexical Sense, both of which were described in the previous section. Presently the Semantic Complexity Level is set only for verbs; for all other syntactic categories Semantic Complexity Level is set to Not Applicable. The other features are specific to the concept's semantic category. Phrases and

clauses also have features, but there wasn't enough space in the figure above to include their popups.

As was mentioned above, figure 4 shows the semantic representation for the sentence *I should finish reading these books*. The popup for the verb shows that the Aspect is Completive and the Mood is 'should'. So those two features generate *should finish*. The Person feature below JOHN is First indicating that John is the speaker. There are no pronouns in the semantic representations, but every noun has a Person feature. The Number feature on BOOK is Trial and the Distance feature is Near Speaker. So those two features generate *these books*. You can now see how that semantic representation produces *I should finish reading these books*. Following is a complete listing of the semantic/syntactic categories, their features and their values. A brief discussion of each feature will also be included.

1) Objects – Nouns

A. Noun List Index: 1, 2, 3, …

The Noun List Index feature is distinct from the other features. Your grammar will never directly access this feature. In other words, you won't write rules that say something like, "When the Noun List Index is 3, add the affix XYZ." The Noun List Index feature is used to determine whether two nominals are the same or different. For example, this feature is used to identify which noun in a relative clause is coreferential with the head noun. This feature is also used to determine if the subject of an object complement is coreferential with the subject of the matrix clause. For example, $JOHN_1$ WANT [$JOHN_1$ BUY CAR_2] represents *John wants to buy a car*. Similarly, $JOHN_1$ WANT [$MARY_2$ BUY CAR_3] represents *John wants Mary to buy a car*. In the first example where the subject noun in the matrix clause has the same index as the subject noun in the complement clause, you'll want a particular set of rules to apply; when the two nouns have a different index as in the second example, you'll want a different set of rules to apply. The Noun List Index feature is also useful when working with a clause chaining language that uses a switch reference system. To properly generate the Same Subject and Different Subject morphemes, you'll use the Noun List Index feature. This feature will be discussed more thoroughly in the tutorial that describes transfer rules.

B. Number: Singular, Dual, Trial, Quadrial, Plural, Paucal

Very few languages morphologically distinguish all six of these possible values. In a subsequent lesson you'll write rules that will collapse feature values so that your grammar only has to deal with the values that are relevant to your language. So if your language only distinguishes Singular and Plural, your grammar will only need to deal with Singular and Plural.

C. Participant Tracking: First Mention, Integration, Routine, Exiting, Restaging, Offstage, Generic, Interrogative, Frame Inferable

A few languages introduce participants onto the discourse stage in a two step process. The first time a participant is mentioned in a discourse, it is tagged as First Mention. The second time it's mentioned, it's tagged as Integration. For the remainder of the discourse, the nominal will be tagged as Routine. The last time a participant appears in a discourse, it will be tagged as Exiting. If a participant is brought back onto the discourse stage, it will be tagged as Restaging. An example of Generic is found in Adjectives 1:8, *John read many books*. Here the nominal BOOK is tagged as Generic. When a nominal is tagged as Interrogative, we generally want the English word *which* before it. An example is found in Clauses 1:98, *Which book did John read?* There are four exceptions to this: when the concepts PERSON, THING, TIME and PLACE are tagged as interrogative, you should generate the equivalent of *who*, *what*, *when* and *where*. Examples are found in Clauses 1:91-94, *Who read a book? What did John read? When did John read a book? Where did John read a book?* When a nominal is tagged as Frame Inferable, English inserts the definite article *the* as in *I bought a car last week. Unfortunately the engine had a problem*. In the second sentence ENGINE would be tagged as Frame Inferable rather than First Mention.

D. Polarity: Affirmative, Negative

Virtually all of the nouns in the semantic representations are tagged as Affirmative. When a noun is tagged as Negative, English places the word *no* before it as in *No man has climbed that mountain*. In this proposition MAN is tagged with Negative Polarity. Your language may form the semantic equivalent using an affirmative generic plural noun with a negative verb as in *Men have not climbed that mountain*.

E. Distance: Not Applicable, Near Speaker and Listener, Near Speaker, Near Listener, Remote within sight, Remote out of sight, Temporally Near, Temporally Remote, Contextually Near, Contextually Remote

This feature is used to generate the demonstratives *this* and *that*. Physical objects will be assigned the values Near Speaker and Listener, Near Speaker, Near Listener, Remote within sight, and Remote out of sight. Only temporal objects will be assigned the values Temporally Near and Temporally Remote. For example, we can talk about *this Christmas* or *that weekend*. The objects that are tagged as Contextually Near or Contextually Remote are those objects that aren't physically present, but are still on the discourse stage. For example, we can talk about *that man* even when his physical location is unknown.

F. Person: First, Second, Third, First & Second, First & Third, Second & Third, First & Second & Third

Many languages have a First Person Inclusive pronoun and a First Person Exclusive pronoun. First Person Inclusive is the equivalent of First & Second. First Person Exclusive is equivalent to First & Third. The values Second & Third and First & Second & Third are included only for completeness; they will probably never occur in the semantic representations.

G. Surface Realization: Not Applicable, Always a Noun, Unambiguous Pronoun

The vast majority of the nouns in the semantic representations are tagged with a Surface Realization value of Not Applicable. When this is the case, your rules will decide which nouns should be realized by target nouns and which should be realized by pronouns. The values Always a Noun and Unambiguous Pronoun are only used with nouns that have a Person value of Third; nouns with a Person value of anything other than Third are always tagged as Not Applicable. This is because languages use pronouns differently, particularly third person singular pronouns. A language such as English uses third person pronouns extensively; a language such as Korean rarely uses third person singular pronouns. So when a noun is tagged as Not Applicable, your rules will determine whether or not a pronoun should be generated. When a noun is tagged as Always a Noun, you should generate a target noun. For example, John 1:1 says *In the beginning was the Word, and the Word was with God, and the Word was God*. In that verse we want all three occurrences of *Word* to be realized with a noun. Therefore those three instances of WORD are tagged as Always a Noun. The value Unambiguous Pronoun is reserved for situations where there are two or more simple propositions referring to the same Third Person Singular nominal. For example, *John went to the store. John bought bread*. The second occurrence of JOHN would be tagged as Unambiguous Pronoun. If your language would use a third person singular pronoun in the second sentence, then you can use the Unambiguous Pronoun value to generate a pronoun.

H. Participant Status: Not Applicable, Protagonist, Antagonist, Major Participant, Minor Participant, Major Prop, Minor Prop, Significant Location, Insignificant Location

A few languages morphologically indicate protagonists and antagonists so they've been marked throughout the semantic representations. It's fairly common for languages to introduce significant participants, props and locations in a different way than the less significant participants, props and locations. For example, when a significant prop or location is first mentioned, it may be placed at the beginning of the clause. If your language doesn't make use of this information, you can hide this feature. Hiding irrelevant features will be discussed in the final tutorial.

2) Events - Verbs

A. Time: Past, Future, Present, Immediate Past, Earlier Today, Yesterday, 2 Days Ago, 3 Days Ago, A Week Ago, A Month Ago, A Year Ago, During Speaker's Lifetime, Historic Past, Eternity Past, Unknown Past, Discourse, Immediate Future, Later Today, Tomorrow, 2 Days from Now, 3 Days from Now, A Week from Now, A Month from Now, A Year from Now, Unknown Future, Timeless

Time is the semantic equivalent of the syntactic term Tense. Many languages have multiple degrees of past tense and multiple degrees of future tense. Since those languages define their degrees differently, the divisions here are finer than the divisions made by any one particular language. Therefore every language will require rules to collapse these Time values into a smaller set. Most of the verbs in the semantic representations are tagged as Discourse; the other values are only used in direct speech. In many languages the type of discourse determines the general tense that is used. English uses past tense for narrative discourse, but that isn't universal. Your rules that generate tense will need to look at the Discourse Type feature that is on every clause.

Timeless tense always coincides with Gnomic aspect. Timeless tense and Gnomic aspect are used in sentences that are always true, e.g., *Oil floats on water*, *The sun rises in the east*, *God loves us*, etc. English encodes timeless tense and gnomic aspect with present tense, but again that isn't a universal.

B. Aspect: Unmarked, Habitual, Imperfective, Progressive, Completive, Inceptive, Cessative, Continuative, Gnomic

The aspects included here are the aspects that significantly affect the meaning of the text. Some of the more subtle aspects such as Perfective have not been included in this list because they don't significantly affect the generated text. Another common aspect, Perfect, has not been included in this list. Indo-European languages generally use Perfect aspect to signal flashback. Flashback is a Salience Band and Salience Bands are indicated on each clause. Another notable aspect that is missing from this list is Iteration. In the semantic representations Iteration is indicated with either an adverbial phrase containing the concept REPEATEDLY, or an oblique NP containing the specific number of times the event was iterated. The vast majority of the verbs in the semantic representations are tagged with Unmarked aspect. Examples of how to handle the other aspects follow:

Habitual – add a word such as *regularly* or *always*

Imperfective – the action is portrayed as ongoing. English uses a participle as in *While John was sleeping …*

Progressive – the action is portrayed as progressing. English again uses a participle as in *John was walking to the store and …*

Completive – add a word such as *finish*

Inceptive – add a word such as *begin* or *start*

Cessative – add a word such as *stop*

Continuative – add a word such as *continue*

C. Mood: Indicative, Definite Potential, Probable Potential, 'might' Potential, Unlikely Potential, Impossible Potential, 'must' Obligation, 'should' Obligation, 'should not' Obligation, Forbidden Obligation, 'may' (permissive)

Mood indicates an event's relationship to reality. Indicative mood signals that the event did or will happen; all of the other moods indicate that the event hasn't actually happened. Examples of how English encodes each of these moods follow:
Definite Potential: *John will definitely walk to the store*.
Probable Potential: *John will probably walk to the store*.
'might' Potential: *John might walk to the store*.
Unlikely Potential: *John will probably not walk to the store*.
Impossible Potential: *John definitely will not walk to the store*.
'must' Obligation: *John must walk to the store*.
'should' Obligation: *John should walk to the store*.
'should not' Obligation: *John should not walk to the store*.
Forbidden Obligation: *John must not walk to the store*.
Permissive 'may': *John may walk to the store*. (For example, John's parents are giving him permission to walk to the store.)

D. Reflexivity: Not Applicable, Reciprocal, Reflexive

Reciprocal and Reflexive situations are marked on the verb even though the morphological encoding for these events generally appears in the object phrase. An example of a reciprocal clause is *We saw each other*. An example of a reflexive clause is *I saw myself*.

E. Polarity: Affirmative, Negative, Emphatic Affirmative, Emphatic Negative

Emphatic Affirmatives are encoded in English with something like *I will definitely read that book*. An example of an Emphatic Negative is *I definitely did not read that book*.

3) Object Attributes - Adjectives

A. Degree: Not Applicable, Comparative, Superlative, Intensified, Extremely Intensified, 'too,' 'overly', 'less', 'least'

Most adjectives in the semantic representations are tagged with a Degree value of Not Applicable. For example, in the sentence *John is tall*, and in the phrase *the tall man*, both occurrences of TALL would be tagged with a Degree value of Not Applicable. English examples for the other values of Degree follow:
Comparative: *John is taller than Steve*.

Superlative: *John is the tallest man in the room*.
Intensified: *John is very tall*.
Extremely Intensified: *John is extremely tall*.
'too,' 'overly': *John is too tall*.
'less': *This book is less important than that book*.
'least': *That book is the least important book*.

Not all languages have comparative constructions as in *John is taller than Steve*. The semantic equivalent in some languages may be *Steve is tall but John is very tall*. If your target language uses the latter construction, you'll be able to write a rule that will restructure all sentences that contain comparative adjectives.

4) Event Attributes - Adverbs

A. Degree: Not Applicable, Comparative, Superlative, Intensified, Extremely Intensified, 'too', 'less', 'least'

Similar to Adjective Degree, most adverbs in the semantic representations are tagged with a Degree value of Not Applicable. For example, in the sentence *John walked quickly*, QUICKLY would be tagged with a Degree value of Not Applicable. English examples of the other values for Degree are listed below.
Comparative: *John walked more quickly than Steve*.
Superlative: *John walked the quickest*.
Intensified: *John walked very quickly*.
Extremely Intensified: *John walked extremely quickly*.
'too': *John walked too quickly*.
''less': *John walked less quickly than Steve*.
'least': *John walked the least quickly*.

Similar to the comparative adjectives, not all languages have comparative adverbs. Rather than saying *John walked more quickly than Steve*, some languages may say *Steve walked quickly, but John walked very quickly*. If that's the case in your language, you'll be able to write a rule to deal with those constructions. The semantic representations will never compare two events that are different. For example, the semantic representations will never say *John walked more quickly than Steve ran*. In that sentence one person's *walking* is being compared with another person's *running*. Although this is permissible in real language, it's not permissible in the semantic representations.

5) Relations - Adpositions

There aren't any features associated with relations.

6) Conjunctions

There aren't any features associated with conjunctions.

7) Particles

There aren't any features associated with particles.

8) Object Phrases - Noun Phrases

A. Type: Simple, First Coordinate, Last Coordinate, Coordinate

In TBTA's semantic representations there are four types of phrases: noun phrases, verb phrases, adjective phrases and adverb phrases. There are no prepositional phrases. Prepositions occur in noun phrases as well as in adverbial clauses. All four phrase types have a feature called Type. In general the Type feature indicates how many phrases there are of a particular type. For noun phrases, the Type feature indicates how many noun phrases with a particular semantic role occur in that clause. For example, if a Patient NP is tagged as Simple, then it is the only Patient NP in that clause. In the sentence *John kicked the ball*, the NP containing BALL would be tagged as Simple because it is the only Patient NP in the clause. If there are multiple patient NPs in a clause, then the first will be tagged with First Coordinate and the last will be tagged with Last Coordinate. If there are three or more NPs in a clause that all have the same semantic role, then the NPs that aren't first or last will all be tagged with Coordinate. So in the sentence *John kicked the ball, the bat, the base and the glove*, the NP containing BALL would be tagged as First Coordinate, the NPs containing BAT and BASE would both be tagged as Coordinate, and the NP containing GLOVE would be tagged as Last Coordinate.

B. Semantic Role: Participant, Patient, State, Source, Destination, Instrument, Addressee, Beneficiary, Not Applicable

Linguists have proposed many different sets of semantic roles[2]. The set of semantic roles used in TBTA has intentionally been kept very small. Rather than using a large set of very descriptive and precise semantic roles, TBTA uses a small set of generic semantic roles. This makes it much easier for users to write rules that will convert the generic semantic roles into roles that are appropriate for their languages.

The term Participant is used as a generic semantic role to cover the roles traditionally labeled as agents, forces, and experiencers. Participants are always the event's external argument. The Patient semantic role generally is used for the most affected argument. The State semantic role is used almost exclusively in stative clauses. For example, in the sentence *John is a teacher*, the NP containing TEACHER is labeled as a State and the NP containing JOHN is labeled as a Participant. The other semantic roles are self explanatory.

Even with a small set of semantic roles, the role assigned to a particular nominal may be debatable. You may not always agree with the role assignments

[2] For additional discussion of semantic roles, see Longacre, Robert E., *The Grammar of Discourse*, Second Edition, p. 153.

in the semantic representations, and in some cases the role assignments may seem somewhat arbitrary. For example, in the sentence *Mary named the baby John* (Noun Phrases 1:25), the NP containing MARY is the Participant, the NP containing BABY is the Patient, and the NP containing JOHN is a State. The precise identification of each nominal's semantic role isn't important. When generating text in another language, what's important is that the semantic representations label the various roles for each event consistently. One of the tools used during the development of the semantic representations guarantees that the semantic roles will be labeled consistently for every event. Therefore you'll be able to write rules that will change the semantic roles to whatever is appropriate for your language.

In the main menu select Lexicon, Source to Target Mappings, Events to Verbs. When you look at the events in the ontology, there's a special column labeled Case Frame. In that column there's a sequence of coded letters. To see what those letters mean, rest your cursor in that column for a particular event. You'll see a popup telling you which semantic roles are obligatory for that event, which are optional, and which never occur. For example, scroll down to ANSWER-A and rest the cursor in its Case Frame cell. The popup will tell you that ANSWER-A always has a Participant NP, it has an optional Patient NP, it never has a State, Source or Instrument NP, it takes optional Destination and Beneficiary NPs, it takes an optional Patient Proposition, and it never takes a Participant Proposition. You'll then see a series of examples illustrating each of the possible situations. The first example is found in Esther 5:7 and illustrates ANSWER-A with just a Participant NP and a Destination NP. That proposition produces the sentence *Esther answered the king*. The next example illustrates ANSWER-A with a Participant NP, a Destination NP, and a Patient Proposition. As you can see in the popup, there's an example for each possible case frame situation for ANSWER-A. You'll see this same information in a popup when you're looking at the semantic representation of a verse and you rest the cursor on a verb. For example, in the main menu select Generator, Lexicon and Grammar Development. Rest the cursor on any verb and you'll see a popup showing its case frame and examples of all its possible argument structures. To see this popup, make sure that you rest the cursor on the verb itself and not on its features.

C. Pragmatic Marking: None, Fronted, Dislocated

This feature is rarely used. If a particular NP was emphasized in the original texts by fronting it, then that information will be marked here. You may then choose how to emphasize that particular NP in your language. Note that your language's method of emphasizing a particular noun phrase may not be by fronting it; instead you should find an appropriate method of emphasizing the NP in your language. An example of a fronted NP in the original texts is found in Genesis 1:1 which says *<u>In the beginning</u> God created the heavens and the earth*. The noun phrase *in the beginning* is tagged as Fronted because that phrase is fronted in the Hebrew text.

Another method of emphasizing a particular nominal in the original texts was to dislocate it. For example, in Genesis 1:27 we read *Male and female, he created them*. Because the original texts emphasized these two noun phrases by dislocating them, they will be tagged as Dislocated in the semantic representations. But as was mentioned in the previous paragraph, this does not mean that you should dislocate these noun phrases in your target language. You should find a natural and appropriate method of emphasizing these nominals in your target text.

9) Event Phrases - Verb Phrases

A. Type: Simple, First Coordinate, Last Coordinate, Coordinate

As was discussed under Noun Phrases, this feature indicates how many verb phrases are present in a particular clause. Since almost all of the clauses have a single verb, it is extremely rare that you'll find a coordinate VP. In fact, in the semantic representations developed to date, there is only one example and it is found in Verbs 1:29. This clause produces *John walked and sang*.

10) Object Attribute Phrases - Adjective Phrases

A. Type: Simple, First Coordinate, Last Coordinate, Coordinate

Coordinate adjective phrases are very rare. You can see an example in Adjectives 1:12 which produces *That book is big and red*. The adjective phrase containing BIG is tagged as First Coordinate and the adjective phrase containing RED is tagged as Last Coordinate.

B. Usage: Attributive, Predicative

English adjectives may be used either attributively, as in the phrase *the tall man*, or they may be used predicatively as in the clause *John is tall*. Adjectives used attributively will always be in an adjective phrase that's embedded in a noun phrase. Adjectives used predicatively will always be in an adjective phrase that's in a clause. An example of an attributive adjective can be found at Adjectives 1:1. An example of a predicative adjective can be found at Adjectives 1:11. When an adjective is used predicatively, the verb will always be BE-D. You may want to write transfer rules that will convert predicative adjectives to target verbs. If that's the case, your transfer rule should be triggered by BE-D. This operation will be described thoroughly in a subsequent tutorial.

11) Event Attribute Phrases - Adverb Phrases

A. Type: Simple, First Coordinate, Last Coordinate, Coordinate

Similar to the adjective phrases, coordinate adverb phrases are very rare. You can see an example in Adverbs 1:6 which produces *John walked quickly and carefully*. The adverb phrase containing QUICKLY is labeled First Coordinate and the adverb phrase containing CAREFULLY is labeled Last Coordinate.

12) Propositions - Clauses

A. Type: Independent, Coordinate Independent, Restrictive Thing Modifier, Descriptive Thing Modifier, Event Modifier, Participant, Patient, Attributive Patient

There are two general categories of clauses: independent and dependent. Independent clauses are able to stand alone and are labeled Independent. Clauses labeled Coordinate Independent are independent clauses that begin with a conjunction. In order to keep the sentences simple, each clause in the semantic representations has only one verb. For example, *John walked to the store. And John bought some bread*. The second sentence is a coordinate independent clause because it's independent and it begins with a conjunction. You'll eventually write rules that will combine simple clauses into more complex clauses when appropriate conditions are met. The other clause types listed above are all dependent clauses. Restrictive Thing Modifier is the semantic name for a restrictive relative clause, e.g., *The man that I saw bought that book*. Descriptive Thing Modifier is the semantic name for a descriptive relative clause, e.g., *John, who will be arriving later today, will open the meeting*. In this sentence the listeners already know who John is; the relative clause adds extra information rather than identifying who John is. Event Modifier is the semantic equivalent of an adverbial clause. All adverbials clauses begin with a relation/preposition that signals the relationship between the main clause and the adverbial clause. A Participant proposition is a subject complement clause, e.g., *It pleased Mary that John read that book*. English generally inserts the expletive *it* and then postposes the subject complement. A Patient proposition is an object complement, e.g., *John wanted Mary to read a book*. An Attributive Patient proposition is an object complement for an adjective, e.g., *John is afraid to read that book*.

B. Illocutionary Force: Declarative, Imperative, Content Interrogative, Yes-No Interrogative, 'may' Imperative, 'let' Imperative

All of these values are self explanatory except the final two. A verse with several 'may' imperatives is Ruth 4:11, *May the Lord bless you …, and may you become powerful …, may you become famous* … You'll need to determine your language's equivalent method for encoding these constructions. The equivalent in your language may be something like "*I pray that the Lord will bless you …, and I pray that you'll become powerful …, and I pray that you'll become famous …*" You'll be able to write rules that will generate the semantic equivalent

of ‘may’ imperatives in your language. ‘let’ imperatives only occur in the first chapter of Genesis. For example, Gen. 1:3, *God said, “Let there be light*.”

C. Topic NP: Participant, Patient, State, Source, Destination, Instrument, Beneficiary

Every clause is marked to signal the semantic role of the topicalized nominal. Generally the topic NP is the Participant; when this is the case, the clause is active. If the clause’s topic NP is the Patient, the clause is passive. Your target language may not have active and passive constructions, but every language has some method of emphasizing one particular nominal over the other nominals. The Topic NP feature indicates which nominal is to be emphasized.

D. Discourse Genre: Climactic Narrative Story, Episodic Narrative Story, Narrative Prophecy, Expository, Behavioral Hortatory, Behavioral Eulogy, Procedural, Persuasive, Expressive, Descriptive, Epistolary, Dramatic Narrative, Dialog

Every clause is marked to indicate its discourse genre. The discourse genres listed above come from *The Grammar of Discourse* by Robert E. Longacre, p. 10.

E. Direct Quote 1: Not Applicable, God to Man, Man to God, God to Woman, Woman to God, God to Angel, Angel to God, God to Satan, Satan to God, Jesus to God the Father, God the Father to Jesus, Jesus to Man, Man to Jesus, Jesus to Woman, Woman to Jesus, Jesus to Disciple, Disciple to Jesus, Jesus to Religious Leader, Religious Leader to Jesus, Jesus to Government Leader, Government Leader to Jesus, Jesus to demon, demon to Jesus, Jesus to Satan, Satan to Jesus, Man to Man, Man to Woman, Woman to Woman, Woman to Man, Husband to Wife, Wife to Husband, Father to Grown Son, Grown Son to Father, Father to Grown Daughter, Grown Daughter to Father, Father to Child, Child to Father, Man to Angel, Angel to Man, Man to Thing, Employer to Hired Worker, Hired Worker to Employer, Owner to Slave, Slave to Owner, Man to Religious Leader, Religious Leader to Man, Man to Government Leader, Government Leader to Man, Man to King, King to Man, Man to Queen, Queen to Man, Mother to Grown Son, Grown Son to Mother, Mother to Grown Daughter, Grown Daughter to Mother, Mother to Child, Child to Mother, Woman to Angel, Angel to Woman, Angel to Angel

Many languages use a system of honorifics when people speak to one another. For example, when parents talk to their children, they may use one set of personal pronouns or a particular verbal affix, but when children talk to their parents, the children use a different set of pronouns or a different verbal affix. Similarly men may have one style of speaking, and women may have another. One culture may honor its priests, another culture may honor its warriors. Since it’s impossible to predict how or when a particular language will indicate honor or

deference during direct speech, every direct quote in the semantic representations has been marked to indicate the general category of the speaker and the listener. For example, the first value listed above is 'God to Man'. In Genesis 2:16 God tells Adam that he is free to eat the fruit in the garden, but Adam must not eat from the tree of knowledge of good and evil. Those sentences are direct quotes and they're tagged as 'God to Man'. The next feature, Direct Quote 2, contains additional categories of speakers and listeners. In the final tutorial you'll learn how to collapse these many different generic categories down into a few specific categories that are relevant for your language. You'll then be able write rules that generate the appropriate pronouns or add the appropriate verbal affix for the particular speech situation.

F. Direct Quote 2: Not Applicable, King to Officials, Officials to King, Queen to Officials, Officials to Queen, Officer to Soldier, Soldier to Officer, Man Praying, Woman Praying, Boy Praying, Girl Praying, Woman to Religious Leader, Religious Leader to Woman, Woman to Government Leader, Government Leader to Woman, Woman to King, King to Woman, Woman to Queen, Queen to Woman, Religious Leader to Religious Leader, Government Leader to Government Leader, Prophet speaking for God to people, Man to Enemy, Woman to Enemy, Boy to Man, Boy to Woman, Girl to Man, Girl to Woman, Man to Boy, Man to Girl, Woman to Boy, Woman to Girl, King to Queen, Queen to King

This feature is a continuation of the previous feature.

G. Salience Band: Not Applicable, Pivotal Storyline, Primary Storyline, Secondary Storyline, Script Predictable Actions, Backgrounded Actions, Flashback, Setting, Irrealis, Evaluation, Cohesive Material

In every discourse regardless of its genre, there is a mainline as well as various types of supportive material. The supporting material encodes progressive degrees of departure from the mainline[3]. Each language has its own techniques for distinguishing the mainline from the different types of supportive material. Additionally each language has its own methods of encoding each of the types of supportive material. For example, English uses perfect aspect to encode Flashback as in the sentence *John had already studied for the test*. Later you'll be able to write rules that apply to one or more particular salience bands.

H. Notional Structure Schema: Not Applicable, Narrative-Exposition, Narrative-Inciting Incident, Narrative-Developing Conflict, Narrative-Climax, Narrative-Denouement, Narrative-Final Suspense, Narrative-Conclusion, Hortatory-Authority Establishment, Hortatory-Problem or Situation, Hortatory-Issuing of Commands, Hortatory-Motivation, Procedural-Problem or Need, Procedural-Preparatory Procedures, Procedural-Main Procedures, Procedural-

[3] For additional discussion of salience bands, see Longacre, Robert E., The Grammar of Discourse, Second Edition, p. 21.

Concluding Procedures, Persuasive-Problem or Question, Persuasive-Proposed Solution or Answer, Persuasive-Supporting Argumentation, Persuasive-Appeal, Expository-Problem or Situation, Expository-Solution or Answer, Expository-Supporting Argumentation, Expository-Evaluation of Solutions

The most significant value of Notional Structure Schema is Narrative Climax. Many languages have a special set of rules that apply only at the peak of a narrative discourse. Therefore each sentence is tagged to indicate its notional structure schema[4].

2.3 Overview of the Structural System used in the Semantic Representations

The structural system in TBTA is generally very simple and clear. As you saw in figure 1, all phrases are enclosed in single brackets, and all clauses are enclosed in double brackets. Noun phrases and adjective phrases may be embedded in noun phrases. Relative clauses are also embedded in noun phrases as shown below in figure 5.

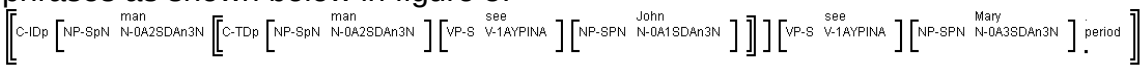

Figure 5. Semantic Representation of *The man that saw John saw Mary*.

All clauses, including subordinate clauses, are enclosed with double brackets. In order to signal the subordination, the double brackets around subordinate clauses are slightly smaller than the double brackets that enclose independent clauses. If you look closely in figure 5, you'll see that the brackets around the relative clause are slightly smaller than the brackets around the main clause.

Shown below in figure 6 is a sentence that contains an object complement clause. Object complements are subordinate clauses, so they have the smaller double brackets, but they're not embedded in any type of phrase.

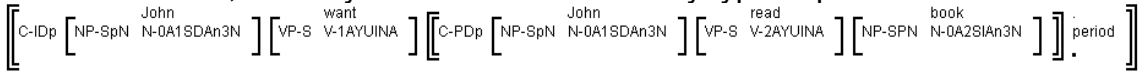

Figure 6. Semantic Representation of *John wanted to read a book*.

Figure 7 shows a sentence with an adverbial clause. All adverbial clauses begin with a preposition that signals the semantic relationship between the main clause and the adverbial clause.

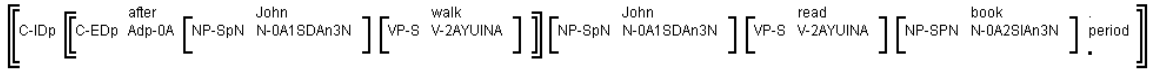

Figure 7. Semantic representation of *After John walked, he read a book*.

Figure 8 contains an attributive patient clause. Notice that it's a subordinate clause embedded in an adjective phrase.

[4] For additional discussion of notional structure schema, see Longacre, Robert E., The Grammar of Discourse, Second Edition, chapters 1 and 2.

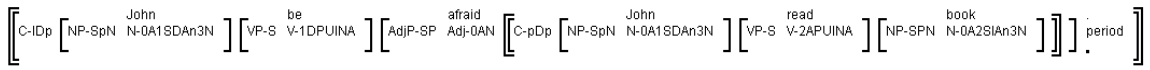

Figure 8. Semantic Representation of *John is afraid to read the book*.

You'll see many more examples of all these structures as you work through the tutorials. The structure of each sentence can be seen a little more clearly if you select the tree view option in the Lexicon and Grammar Development screen. In the main menu select Generator, Lexicon and Grammar Development. Then click the Reference button in the upper left corner and select English Grammar Introduction, Clauses 1:40. You'll see the semantic representation shown above in figure 8. Now click the Show Tree option in the upper right corner. You'll see a flat tree that makes the structure of the sentence a little clearer. Now uncheck that option and you'll see the standard representation. You may select whichever view you prefer. However, many of the rules will contain structures, and in the rules you'll always use the linear view. So you should probably use the linear view rather than the tree view when displaying the semantic representations.

3.0 Conclusion

You've now been introduced to the concepts, features and structures that have been used to build the semantic representations for TBTA. These semantic representations contain a great deal of information; no language will use all of this information. As was mentioned above, the semantic representations reflect an English perspective. They use concepts that have been lexicalized by English, and the structures reflect English sentence structure. Subsequent tutorials will teach you how to restructure these semantic representations so that they become appropriate underlying representations for your language. You'll be able to restructure the representations so that they contain words, features and structures that reflect your language rather than English. The remainder of these tutorials will teach you how to restructure the semantic representations and then synthesize the final target forms. The result will be target text that is easily understandable, grammatically perfect, and semantically equivalent to the semantic representations.